(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 March 2002 (14.03.2002)

PCT

(10) International Publication Number
WO 02/20765 A2

(51) International Patent Classification[7]: C12N 15/00

(21) International Application Number: PCT/US01/28337

(22) International Filing Date:
10 September 2001 (10.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/231,089 8 September 2000 (08.09.2000) US

(71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventor: GLUCKSMANN, Maria, Alexandra; 33 Summit Road, Lexington, MA 02173 (US).

(74) Agents: COLBY, Gary, D. et al.; Akin, Gump, Strauss, Hauer & Feld, L.L.P., Suite 2000, One Commerce Square, 2005 Market Street, Philadelphia, PA 19103-7986 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/20765 A2

(54) Title: 38646, A NOVEL GUANINE NUCLEOTIDE EXCHANGE FACTOR AND USES THEREFOR

(57) Abstract: The invention provides isolated nucleic acids molecules, designated 28646 nucleic acid molecules, which encode a novel guanine-nucleotide exchange factor. The invention also provides also provides antisense nucleic acid molecules, recombinant expression vectors containing 38646 nucleic acid molecules, host cells into which the expression vectors have been introduced, and non-human transgenic animals in which a 38646 gene has been introduced or disruped. The invention still further provides isolated 38646 proteins, fusion proteins, antigenic peptides and anti-38646 antibodies. Diagnostic methods utilizing compositions of the invention are also provided. 38646 expression and activity can be modulated to affect cell shape, motility, cytoskeleton organization, and intracellular protein and vesicle localization or to affect the tensile strength or integrity of a tissue.

38646, A Novel Guanine Nucleotide Exchange Factor And Uses Therefor

BACKGROUND OF THE INVENTION

The Rho family of GTP-binding proteins (including protein subfamilies designated Rho, Rac, and Cdc42) is involved in numerous signaling pathways which affect cytoskeletal organization. These proteins, which exhibit GTPase activity, are also known to be involved in activation of Jun amino-terminal kinase, regulation of p70 S6 kinase, and induction of transcriptional activation of genes containing the serum response element.

The actin cytoskeleton is essential for many cellular functions. For example, the cytoskeleton is involved in maintenance of cell shape, in cell adhesion, in cell motility (e.g., formation of filopodia, lamellipodia, and membrane ruffles). Each of these functions are among those regulated during development of an embryo and in reassembly or remodeling of tissues in adult organisms (e.g., during wound healing, tumor infiltration and metastasis, and immune cell infiltration of tissues).

The actin cytoskeleton is also known to be involved in growth and division of cells. Interaction of chromosomes with actin can regulate their segregation during mitosis and meiosis. In addition, the cytoskeleton is known to be involved in progression of the cell through the cell replication cycle.

Interaction between Rho GTP-binding proteins and the cytoskeleton requires activation of the Rho proteins. This activation involves displacement of GDP bound with an inactive Rho protein by GTP, yielding the corresponding activated Rho protein. Displacement of GDP by GTP can be catalyzed by proteins designated guanine-nucleotide exchange factors (GEFs), of which the Dbl oncoprotein has been proposed as a prototype (Cerione et al., 1996, Curr. Opin. Cell Biol. 8:216-222). Several GEFs have been isolated from various organisms, including the rat frabin gene (Obaishi et al., 1998, J. Biol. Chem. 273:18697-18700), the human fgd1 gene (Zheng et al., 1996, J. Biol. Chem. 271:33169-33172), the human vav gene (Olson et

al., 1996, Curr. Biol. 6:1628-1633), murine and human fgd2 genes (Pasteris et al., 1999, Genomics 60:57-66), and murine fgd3 gene (Pasteris et al., 1994, Cell 79:669-678).

Aberrance or deficiency in GEF expression has been correlated with occurrence of developmental disorders (e.g., faciogenital dysplasia) and cancer. Furthermore, the demonstrated abilities of at least some GEFs (e.g., rat Frabin protein) to interact with both the actin cytoskeleton and with Rho proteins indicates that GEFs are involved in regulation of normal and morbid cytoskeletal processes.

Although a few GEFs, including a few human GEFs, are known, it is likely that other GEFs have yet to be characterized. A significant need exists for identification and characterization of additional GEF proteins and the genes encoding them. The present invention satisfies this need, at least in part, by identifying and characterizing a GEF that is herein designated "38646."

BRIEF SUMMARY OF THE INVENTION

The present invention is based, in part, on the discovery of a novel gene encoding a GEF, the gene being referred to herein as "38646". The nucleotide sequence of a cDNA encoding 38646 is shown in SEQ ID NO: 1, and the amino acid sequence of a 38646 polypeptide is shown in SEQ ID NO: 2. In addition, the nucleotide sequence of the coding region is depicted in SEQ ID NO: 3.

Accordingly, in one aspect, the invention features a nucleic acid molecule that encodes a 38646 protein or polypeptide, e.g., a biologically active portion of the 38646 protein. In a preferred embodiment the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence SEQ ID NO: 2. In other embodiments, the invention provides isolated 38646 nucleic acid molecules having the nucleotide sequence of one of SEQ ID NO: 1 and SEQ ID NO: 3. In still other embodiments, the invention provides nucleic acid molecules that have sequences that are substantially identical (e.g., naturally occurring allelic variants) to the nucleotide sequence of one of SEQ ID NO: 1 and SEQ ID NO: 3. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under stringent hybridization conditions with a nucleic acid molecule having a sequence comprising the nucleotide sequence

of one of SEQ ID NO: 1 and SEQ ID NO: 3, wherein the nucleic acid encodes a full length 38646 protein or an active fragment thereof.

In a related aspect, the invention further provides nucleic acid constructs that include a 38646 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included are vectors and host cells containing the 38646 nucleic acid molecules of the invention, e.g., vectors and host cells suitable for producing 38646 nucleic acid molecules and polypeptides.

In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for detection of 38646-encoding nucleic acids.

In still another related aspect, isolated nucleic acid molecules that are antisense to a 38646-encoding nucleic acid molecule are provided.

In another aspect, the invention features 38646 polypeptides, and biologically active or antigenic fragments thereof that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of 38646-mediated or related disorders (e.g., GEF-mediated disorders such as those described herein). In another embodiment, the invention provides 38646 polypeptides having Rho GTP-binding protein activating activity. Preferred polypeptides are 38646 proteins including at least one PH domain and one RhoGEF domain, and preferably having a 38646 activity, e.g., a 38646 activity as described herein. Preferred polypeptides are 38646 proteins including at least one RhoGEF domain, at least two PH domains, and at least one FYVE domain.

In other embodiments, the invention provides 38646 polypeptides, e.g., a 38646 polypeptide having the amino acid sequence shown in SEQ ID NO: 2;; an amino acid sequence that is substantially identical to the amino acid sequence shown in SEQ ID NO: 2; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO: 1 and SEQ ID NO: 3, wherein the nucleic acid encodes a full length 38646 protein or an active fragment thereof.

In a related aspect, the invention further provides nucleic acid constructs that include a 38646 nucleic acid molecule described herein.

In a related aspect, the invention provides 38646 polypeptides or fragments operatively linked to non-38646 polypeptides to form fusion proteins.

In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably, specifically bind, 38646 polypeptides.

In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 38646 polypeptides or nucleic acids.

In still another aspect, the invention provides a process for modulating 38646 polypeptide or nucleic acid expression or activity, e.g., using the screened compounds. In certain embodiments, the methods involve treatment of conditions related to aberrant activity or expression of the 38646 polypeptides or nucleic acids, such as conditions involving aberrant or deficient cell shape, growth, differentiation, adhesion, or motility, conditions involving aberrant or deficient development (i.e., of an organism or an organ), and conditions involving aberrant or deficient progression of a cell through the cell cycle.

The invention also provides assays for determining the activity of or the presence or absence of 38646 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

In further aspect the invention provides assays for determining the presence or absence of a genetic alteration in a 38646 polypeptide or nucleic acid molecule, including for disease diagnosis.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

Figure 1 depicts a cDNA sequence (SEQ ID NO: 1) and predicted amino acid sequence (SEQ ID NO: 2) of human 38646. The methionine-initiated open reading frame of human 38646 (without the 5'- and 3'-non-translated regions) starts at nucleotide 115 of SEQ ID NO: 1, and the coding region (not including the terminator codon; shown in SEQ ID NO: 3) extends through nucleotide 2413 of SEQ ID NO: 1.

Figure 2 depicts a hydropathy plot of human 38646. Relatively hydrophobic residues are shown above the dashed horizontal line, and relative hydrophilic residues are

below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines below the hydropathy trace. The numbers corresponding to the amino acid sequence of human 38646 are indicated. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, i.e., a sequence above the dashed line, e.g., the sequence of about residues 440-460 of SEQ ID NO: 2; all or part of a hydrophilic sequence, i.e., a sequence below the dashed line, e.g., the sequence of residues 168-185 or 190-210 of SEQ ID NO: 2; a sequence which includes a cysteine residue; or a glycosylation site.

Figure 3, comprising Figures 3A and 3B, is a manual alignment of the amino acid sequences or rat Frabin protein (SEQ ID NO: 4; "Frabin"; Obaishi et al., 1998, J. Biol. Chem. 273:18697-18700; GenBank™ accession number AF038388), and 38646 protein (SEQ ID NO: 2), in which identical residues are indicated by ":".

Figure 4, comprising Figures 4A through 4F, is an alignment of the amino acid sequences of murine FGD1 protein (SEQ ID NO: 5; "mFGD1"; GenBank™ accession number NP_032027), human FGD1 protein (SEQ ID NO: 6; "hFGD1"; GenBank™ accession number NP_004454), murine FGD3 protein (SEQ ID NO: 7; "mFGD3"; GenBank™ accession number NP_056574), rat frabin protein (SEQ ID NO: 4; "rFrabin;" GenBank™ accession number AF038388), murine FGD2 protein (SEQ ID NO: 8; "mFGD2"; GenBank™ accession number NP_038738), and 38646 protein (SEQ ID NO: 2). The mFGD1, hFGD1, mFGD3, rFrabin, and mFGD2 sequences were aligned using the CLUSTALW multiple sequence alignment program. For those sequences, an asterisk ("*") in the sixth line of the alignment indicates that the same residue was present at that position in all five sequences, a colon (":") indicates that the residues present at that position were highly conservative alternative residues, and a period (".") indicates that the residues present at that position were less conservative alternative residues. The 38646 sequence was manually aligned beneath the CLUSTALW alignment.

Figure 5 is an alignment of a portion of the amino acid sequence of 38646 protein (i.e., residues 12 to 185 of SEQ ID NO: 2; "38646") with a portion (SEQ ID NO: 9; "878") of the amino acid sequence disclosed as sequence number 878 in PCT publication number WO 01/12659.

DETAILED DESCRIPTION OF THE INVENTION

The human 38646 cDNA sequence (Figure 1; SEQ ID NO: 1), which is approximately 2561 nucleotide residues long including non-translated regions, contains a predicted methionine-initiated coding sequence of about 2298 nucleotide residues, excluding termination codon (i.e., nucleotide residues 115-2413 of SEQ ID NO: 1; also shown in SEQ ID NO: 3). The coding sequence encodes a 766 amino acid residue protein having the amino acid sequence SEQ ID NO: 2.

Human 38646 contains the following regions or other structural features: a predicted RhoGEF domain (PF00621) at about amino acid residues 210 to 392 of SEQ ID NO: 2, an FYVE domain at about amino acid residues 554 to 620 of SEQ ID NO: 2, and two predicted PH domains at about amino acid residues 423 to 521 and 644 to 740 of SEQ ID NO: 2.

The human 38646 protein has predicted N-glycosylation sites (Pfam accession number PS00001) at about amino acid residues 35-38, 77-80, 150-153, 256-259, 315-318, and 436-439 of SEQ ID NO: 2; predicted cAMP-/cGMP-dependent protein kinase phosphorylation sites (Pfam accession number PS00004) at about amino acid residues 40-43 and 761-764 of SEQ ID NO: 2; predicted protein kinase C phosphorylation sites (Pfam accession number PS00005) at about amino acid residues 79-81, 155-157, 217-219, 282-284, 317-319, 466-468, 493-495, 529-531, 576-578, and 708-710 of SEQ ID NO: 2; predicted casein kinase II phosphorylation sites (Pfam accession number PS00006) located at about amino acid residues 25-28, 186-189, 325-328, 438-441, 481-484, 493-496, 505-508, 545-548, 622-625, and 753-756 of SEQ ID NO: 2; predicted N-myristoylation sites (Pfam accession number PS00008) at about amino acid residues 30-35, 84-89, 92-97, 243-248, 462-467, and 604-609 of SEQ ID NO: 2; and a predicted amidation site (Pfam accession number PS00009) at about amino acid residues 549-552 of SEQ ID NO: 2.

For general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer et al. (1997, Protein 28:405-420).

The 38646 protein contains a significant number of structural characteristics in common with members of the GEF family. The term "family" when referring to the protein and nucleic acid molecules of the invention means two or more proteins or nucleic acid

molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin, or alternatively, can contain homologues of non-human origin, e.g., GEF proteins for any species described in the art (e.g., those described in references cited therein). Members of a family can also have common functional characteristics. For example, many GEF family members are involved in determining cytoskeleton organization and intracellular localization. GEF family members have been shown to induce generation and growth of cellular extensions (e.g., actin microspikes or filopodia) and alter cell shape and motility.

A 38646 polypeptide can include a RhoGEF domain. As used herein, the term " RhoGEF domain" refers to a protein domain having an amino acid sequence of about 100-300 amino acid residues in length, preferably, at least about 150-250 amino acid residues, more preferably about 175-225 amino acid residues, even more preferably about 207 amino acid residues or about 182 amino acid residues and has a bit score for the alignment of the sequence to the RhoGEF domain (HMM) of at least 50 or greater, preferably 60 or greater, more preferably, 75 or greater, and most preferably, 100 or greater. The RhoGEF domain has been assigned the PFAM accession PF00621.

In a preferred embodiment, 38646 polypeptide or protein has a RhoGEF domain or a region which includes at least about 100-300, more preferably about 150-250, 175-1225, 207 or 182 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a RhoGEF domain, e.g., the RhoGEF domain of human 38646 (e.g., residues 210-392 of SEQ ID NO: 2).

A 38646 polypeptide can also include a PH domain (i.e., a pleckstrin homology domain). As used herein, the term "PH domain" refers to a protein domain having an amino acid sequence of about 50-150 amino acid residues in length, preferably, at least about 75-100 amino acid residues, more preferably about 85, 96, or 98 amino acid residues and has a bit score for the alignment of the sequence to the PH domain (HMM) of at least 50 or greater, preferably 60 or greater, more preferably, 75 or greater, and most preferably, 100 or greater. The PH domain has been assigned the PFAM accession PF00169.

In a preferred embodiment, 38646 polypeptide or protein has a PH domain or a region which includes at least about 50-150, more preferably about 75-100, 85, 96, or 98 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a PH domain, e.g., a PH domain of human 38646 (e.g., either the domain at residues 423-521 or the domain at residues 644-740 of SEQ ID NO: 2).

A 38646 polypeptide can also include an FYVE domain (i.e., a FYVE zinc finger domain; a domain comprising the sequence FYVE). As used herein, the term "FYVE domain" refers to a protein domain having an amino acid sequence of about 50-150 amino acid residues in length, preferably at least about 60-80 amino acid residues, more preferably about 72 or 66 amino acid residues and has a bit score for the alignment of the sequence to the FYVE domain (HMM) of at least 50 or greater, preferably 60 or greater, more preferably, 75 or greater, and most preferably, 100 or greater. The FYVE domain has been assigned the PFAM accession PF01363.

In a preferred embodiment, 38646 polypeptide or protein has an FYVE domain or a region which includes at least about 50-150, more preferably about 60-80, 72, or 66 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with an FYVE domain, e.g., the FYVE domain of human 38646 (e.g., the domain at residues 554-620 of SEQ ID NO: 2).

To identify the presence of a domain in a 38646 protein, the amino acid sequence of the protein is searched against a database of HMMs (e.g., the Pfam database, release 2.1) using default parameters set forth at the World Wide Web site of Washington University (St. Louis) or other default parameters for domain identification. For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program and score of 15 is the default threshold score for determining a hit. For example, using ORFAnalyzer software, the domains described above were identified in the amino acid sequence of SEQ ID NO: 2. Accordingly, a 38646 protein having at least about 60-70%, more preferably about 70-80%, or about 80-90% homology with a domain described herein for human 38646 is within the scope of the invention.

In one embodiment of the invention, a 38646 polypeptide includes at least one RhoGEF domain. In another embodiment, the 38646 polypeptide includes at least one

RhoGEF domain and at least one PH domain. In another embodiment, the 38646 polypeptide comprises at least one RhoGEF domain, at least two PH domains, and at least one FYVE domain.

The 38646 molecules of the present invention can further include one or more of the N-glycosylation, cAMP-/cGMP-dependent protein kinase phosphorylation, protein kinase C phosphorylation, casein kinase II phosphorylation, N-myristoylation, and amidation sites described herein, and preferably comprises most or all of them.

Because the 38646 polypeptides of the invention can modulate 38646-mediated activities, they can be used to develop novel diagnostic and therapeutic agents for 38646-mediated or related disorders, as described below.

As used herein, a "38646 activity," "biological activity of 38646," or "functional activity of 38646," refers to an activity exerted by a 38646 protein, polypeptide or nucleic acid molecule on, for example, a 38646-responsive cell or on a 38646 substrate (e.g., a protein substrate) as determined in vivo or in vitro. In one embodiment, a 38646 activity is a direct activity, such as association with a 38646 target molecule or activation of a Rho GTP-binding protein. A "target molecule" or "binding partner" of a 38646 protein is a molecule with which the 38646 protein binds or interacts in nature. In an exemplary embodiment, such a target molecule is a 38646 receptor. A 38646 activity can also be an indirect activity, such as a cellular signaling activity mediated by interaction of the 38646 protein with a 38646 receptor (e.g., signaling mediated by activation of a Rho GTP-binding protein, such as signaling relating to cell shape, motility, tissue infiltration, metastasis, differentiation, or procession through the cell cycle).

As indicated in Figure 4, 38646 protein exhibits significant amino acid sequence homology with several known GEF proteins. The amino acid sequence (SEQ ID NO: 2) of 38646 protein also exhibits 170 consecutive residues that are identical to a portion of an amino acid sequence encoded by a cDNA molecule disclosed in PCT publication number WO 01/12659, as indicated in Figure 5.

The 38646 molecules of the present invention are predicted to have similar biological activities as GEF family members. For example, the 38646 proteins of the present invention can have one or more of the following activities:

(1) catalyzing activation of Rho GTP-binding proteins, e.g., by displacing GDP bound with the inactive form of the Rho protein with GTP;

(2) modulating cell shape;

(3) modulating cell growth;

(4) modulating cell deformation;

(5) modulating cell differentiation;

(6) modulating cell adhesion;

(7) modulating infiltration of a cell into or through a tissue;

(8) modulating progression of a cell through the cell cycle;

(9) modulating embryonic development

(10) modulating angiogenesis;

(11) modulating neurite growth;

(12) modulating tumor cell invasion or metastasis;

(13) modulating tumorigenesis;

(14) modulating wound healing;

(15) modulating bone remodeling

(16) modulating apoptosis; and

(17) modulating the tensile strength of a solid tissue.

Thus, 38646 molecules described herein can act as novel diagnostic targets and therapeutic agents for prognosticating, diagnosing, preventing, inhibiting, alleviating, or curing GEF-related disorders. Based on the above-referenced activities, GEF-related disorders can include numerous disorders including developmental disorders (e.g., faciogenital dysplasia, congenital heart diseases, neonatal gastrointestinal, musculoskeletal, neurological, eye, and renal and genitourinary defects), cancers (i.e., including tumorigenesis, tumor growth and spread, and metastasis), bacterial, fungal, yeast, and parasitic infections, autoimmune disorders, muscle and connective tissue disorders, cardiovascular disorders (e.g., heart failure, coronary artery disease, and atherosclerosis), fibrotic diseases (e.g., liver fibrosis), angiogenic disorders (e.g., provision of blood supply to forming or developing tumors or aberrant formation or growth of blood vessels) and bone disorders (e.g., osteoporosis and repair of broken bones).

The expression data disclosed herein confirm that 38646 protein is expressed in a variety of tissues in which the above-referenced activities occur. For example, 38646 is expressed at high levels in normal brain cortex and normal brain hypothalamus, at a lower level in normal spinal cord, and at lower levels in nerve tissue and dorsal root ganglia. These tissues are comprised of cells that are highly differentiated, exhibit specific cell shape characteristics, and require an extensive amount of intracellular organization and intracellular movement. Each of these functions can be regulated by GEF family members such as 38646. Compounds which modulate 38646 expression or 38646 activity can therefore be used to influence the ability of neural cells to differentiate, change shape (e.g., extend dendrites or form synapses), or alter intracellular localization of cellular components (e.g., synaptic vesicles and signaling proteins).

Compounds which modulate the activity of 38646 protein or expression of the gene that encodes it can be used to modulate the ability of CNS cells to form interconnections. These compounds can also be used to repopulate areas of CNS extracellular matrix that have become depopulated owing to disease, injury, or the normal aging process. For example, ischemic injury such as stroke leads to localized death of CNS cells in a brain region to which normal blood supply is inhibited or interrupted. Following a stroke, CNS cells normally do not repopulate the affected area, and neural, sensory, cognitive, and motor defects can result from the loss of brain cells. One factor that can inhibit re-population of the affected area is the relative inability of CNS cells to migrate through CNS extracellular matrix. GEF proteins, such as 38646, induce cell shape and motility changes such as microspike formation and dendrite extension. Enhancing expression or activity of 38646 of CNS cells increases the ability of the cells to move or extend into or through the CNS matrix. Movement of CNS cells into the affected area and reformation or repair of cell-cell interconnections which extend through the affected area can ameliorate or reverse the neural deficit experienced by a stroke victim. Expression, activity, or both, of 38646 can be enhanced ex vivo in CNS cells prior to providing the treated CNS cells to an area of the brain that has experienced ischemic injury.

Other tissues in which intracellular protein and vesicle localization is important include glandular tissues (e.g., pancreas, liver, pituitary gland, and adrenal gland). Expression of 38646 in pancreas is relatively high. Less expression of 38646 was detected in liver and pituitary gland, and lower but significant expression of 38646 occurs in adrenal gland. These

data indicate a role for 38646 activity in modulating glandular secretion by affecting cytoskeletal organization and intracellular localization of secretory proteins and vesicles. Modulation of 38646 expression, activity, or both is useful for treating disorders characterized by aberrant secretion from glands such as the pancreas, and the pituitary and adrenal glands.

Cytoskeleton organization and intracellular protein localization is critical and highly regulated in contractile tissue, such as skeletal muscle and heart. 38646 is highly expressed in skeletal muscle tissue and is implicated in regulation of these activities in skeletal muscle cells (i.e., myocytes). Furthermore, muscle repair after damage (e.g., inflicted by extreme muscle use and fatigue or by surgery or trauma) can be affected by GEF family members such as 38646. During muscle repair, alterations in myocyte motility and shape, modulated by 38646, can affect the rate and extent of repair of damaged muscle tissue.

38646 is expressed in heart, human umbilical vein endothelial cells (HUVEC), and arteries. These tissues are consistently exposed to shear stresses from blood flow and cell shape maintenance, regulated at least in part by 38646 molecules, is important for proper vascular function. Modulation of 38646 expression, activity, or both is useful for modulating vascular tissue integrity and proper vascular function.

38646 is expressed in kidney tissue, indicating a role for 38646 activity in this tissue. Podocytes are kidney cells that exhibit many primary and secondary cell processes or extensions. These cell extensions contain numerous cytoskeletal components such as microfilaments and microtubules. These podocyte processes have an important role in proper kidney filtration and function. Modulation of 38646 expression, activity, or both, can increase or decrease kidney function by altering the size and number of podocyte cellular extensions.

The data disclosed herein regarding expression of 38646 in selected human tissues indicates that 38646 is more highly expressed in cell types that occur in tissues which normally experience significant tensile forces in vivo. Those data, analyzed in conjunction with the other properties of GEF proteins, indicate that 38646 protein modulates the tenacity with which cells bind with one another, with their extracellular matrix, or both. By way of example, 38646 is relatively highly expressed in tissues (e.g., skeletal muscle, artery, and heart tissues) which normally experience high tensile strength and in tissues (e.g., kidney, spinal cord, brain cortex, hypothalamus, and ovary tissues) that normally exhibit a relatively rigid physical

structure. 38646 is expressed at relatively low levels in non-solid tissues (e.g., synovium, peripheral blood mononucleocyte, and leukocyte tissues) and in solid tissues that exhibit relatively low rigidity or are not normally exposed to high tensile forces (e.g., spleen, prostate, lymph node, vein, and bladder tissues).

These observations indicate that 38646 is more highly expressed in tissues in which relatively tight binding occurs between cells or between cells and their extracellular matrix than in cells which bind less tenaciously (or which substantially do not bind) with one another or with extracellular matrix. Although it is possible that 38646 directly mediates binding of cells with surrounding cells or matrix, it is more likely that 38646 instead exerts its effects on cell binding by modulating expression of genes encoding proteins which more directly influence cell binding. By way of example, 38646 can modulate expression of genes which encode proteins which link a cell with another cell or a matrix in its environment or genes which encode proteins that catalyze activation (or de-inhibition) of other proteins more directly involved in cell binding.

Modulating expression or activity of 38646 can affect the tensile strength of a tissue or binding affinity of a cell in which the modulation occurs. Enhancing 38646 expression or activity can enhance intercellular binding, increasing the tensile strength and consistency of a tissue. Inhibiting 38646 expression or activity can inhibit intercellular binding, increasing the likelihood that portions (e.g., individual cells or clumps of cells) of the tissue will separate from the main tissue mass (e.g., the likelihood that a tumor will exhibit metastasis). Modulation of 38646 expression or activity can modulate the relative strength of cell binding in normal and diseased tissues.

Other activities, as described below, include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which 38646 molecules are expressed. Thus, the 38646 molecules can act as novel diagnostic targets and therapeutic agents for controlling disorders involving aberrant activities of these cells.

The 38646 molecules can also act as novel diagnostic targets and therapeutic agents for controlling cellular proliferative and/or differentiative disorders (e.g., hematopoietic neoplastic disorders, carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic

disorders, e.g., leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast, and liver origin.

As used herein, the terms "cancer," "hyperproliferative" and "neoplastic" refer to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. Hyperproliferative and neoplastic disease states can be categorized as pathologic, i.e., characterizing or constituting a disease state, or can be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. "Pathologic hyperproliferative" cells occur in disease states characterized by malignant tumor growth. Examples of non-pathologic hyperproliferative cells include proliferation of cells associated with wound repair.

The terms "cancer" or "neoplasms" include malignancies of the various organ systems, such as affecting lung, breast, ovary, thyroid, pancreas, brain, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus.

The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, ovarian carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term also includes carcinosarcomas, e.g., which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures.

The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, e.g., arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. The disorders can arise from poorly differentiated acute leukemias, e.g., erythroblastic leukemia and acute megakaryoblastic leukemia. Exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML; reviewed in Vaickus, 1991, Crit. Rev. Oncol./Hemotol. 11:267-297); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

The 38646 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO: 2 thereof are collectively referred to as "polypeptides or proteins of the invention" or "38646 polypeptides or proteins". Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as "nucleic acids of the invention" or "38646 nucleic acids." 38646 molecules refer to 38646 nucleic acids, polypeptides, and antibodies.

As used herein, the term "nucleic acid molecule" includes DNA molecules (e.g., a cDNA or genomic DNA) and RNA molecules (e.g., an mRNA) and analogs of the DNA or RNA generated, e.g., by the use of nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated or purified nucleic acid molecule" includes nucleic acid molecules that are separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules that are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5'- and/or 3'-ends

of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kilobases, 4 kilobases, 3 kilobases, 2 kilobases, 1 kilobase, 0.5 kilobase or 0.1 kilobase of 5'- and/or 3'-nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

As used herein, the term "hybridizes under stringent conditions" describes conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in available references (e.g., Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 1989, 6.3.1-6.3.6). Aqueous and non-aqueous methods are described in that reference and either can be used. A preferred example of stringent hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2× SSC, 0.1% (w/v) SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2× SSC, 0.1% (w/v) SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2× SSC, 0.1% (w/v) SDS at 60°C. Preferably, stringent hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2× SSC, 0.1% (w/v) SDS at 65°C. Particularly preferred stringency conditions (and the conditions that should be used if the practitioner is uncertain about what conditions should be applied to determine if a molecule is within a hybridization limitation of the invention) are 0.5 molar sodium phosphate, 7% (w/v) SDS at 65°C, followed by one or more washes at 0.2× SSC, 1% (w/v) SDS at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO: 1 or SEQ ID NO: 3, corresponds to a naturally-occurring nucleic acid molecule.

As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a 38646 protein, preferably a mammalian 38646 protein, and can further include non-coding regulatory sequences and introns.

An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of 38646 protein having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-38646 protein (also referred to herein as a "contaminating protein"), or of chemical precursors or non-38646 chemicals. When the 38646 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 38646 (e.g., the sequence of SEQ ID NO: 1 or SEQ ID NO: 3) without abolishing or, more preferably, without substantially altering a biological activity, whereas an "essential" amino acid residue results in such a change. For example, amino acid residues that are conserved among the polypeptides of the present invention, e.g., those present in the RhoGEF domain are predicted to be particularly non-amenable to alteration, and the residues of the RhoGEF, PH, and FYVE domains of 38646 are also preferably not altered.

A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic

acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), non-polar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 38646 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 38646 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 38646 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO: 1 or SEQ ID NO: 3, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

As used herein, a "biologically active portion" of a 38646 protein includes a fragment of a 38646 protein that participates in an interaction between a 38646 molecule and a non-38646 molecule. Biologically active portions of a 38646 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 38646 protein, e.g., the amino acid sequence shown in SEQ ID NO: 2, which include less amino acids than the full length 38646 proteins, and exhibit at least one activity of a 38646 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 38646 protein, e.g., a domain or motif capable of catalyzing an activity described herein.

A biologically active portion of a 38646 protein can be a polypeptide that for example, 10, 25, 50, 100, 200, 300, or 400 or more amino acids in length. Biologically active portions of a 38646 protein can be used as targets for developing agents that modulate a 38646-mediated activity, e.g., a biological activity described herein.

Calculations of homology or sequence identity between sequences (the terms are used interchangeably herein) are performed as follows.

To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison

purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence (e.g., when aligning a second sequence to the 38646 amino acid sequence of SEQ ID NO: 2 having 86 amino acid residues, preferably at least 100, and more preferably at least 171, 175, 200, 225, 250, 300, 350, 400, 500, 600, 700, or 766 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences. Preferably, a 38646 polypeptide has a sequence that comprises at least 171, 175, 200, 225, 250, 300, 350, 400, 500, 600, 700, or 766 consecutive amino acid residues of SEQ ID NO: 2, or at least 50, 75, 100, 150, 200, 300, 400, 500, or 600 consecutive amino acid residues of residues 182-766 of SEQ ID NO: 2. Also preferably, a 38646 polypeptide exhibits at least one of a function and an epitope of 38646 protein.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman et al. (1970, J. Mol. Biol. 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package (available commercially), using either a BLOSUM 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available commercially), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be

applied to determine if a molecule is within a sequence identity or homology limitation of the invention) are a BLOSUM 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers et al. (1989, CABIOS, 4:11-17) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990, J. Mol. Biol. 215:403-410). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 38646 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 38646 protein molecules of the invention. To obtain gapped alignments for comparison purposes, gapped BLAST can be utilized as described in Altschul et al. (1997, Nucl. Acids Res. 25:3389-3402). When using BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. For example, the BLAST software available at the World Wide Web site of the National Center for Biotechnology Information (National Library of Medicine, National Institutes of Health) can be used.

"Malexpression or aberrant expression," as used herein, refers to a non-wild-type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild-type levels, i.e., over- or under-expression; a pattern of expression that differs from wild-type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild-type) at a predetermined developmental period or stage; a pattern of expression that differs from wild-type in terms of decreased expression (as compared with wild-type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild-type in terms of the splicing size, amino acid sequence, post-transitional modification, or biological activity of the expressed polypeptide; a pattern of expression that

differs from wild-type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild-type) in the presence of an increase or decrease in the strength of the stimulus.

"Subject," as used herein, can refer to a mammal, e.g., a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, e.g., a horse, cow, goat, or other domestic animal.

A "purified preparation of cells," as used herein, refers to, in the case of plant or animal cells, an in vitro preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10%, and more preferably, 50% of the subject cells.

Various aspects of the invention are described in further detail below.

Isolated Nucleic Acid Molecules

In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes a 38646 polypeptide described herein, e.g., a full-length 38646 protein or a fragment thereof, e.g., a biologically active portion of 38646 protein. Also included is a nucleic acid fragment suitable for use as a hybridization probe, which can be used, e.g., to a identify nucleic acid molecule encoding a polypeptide of the invention, 38646 mRNA, and fragments suitable for use as primers, e.g., PCR primers for the amplification or mutation of nucleic acid molecules.

In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO: 1, or a portion of that nucleotide sequence. In one embodiment, the nucleic acid molecule includes sequences encoding the human 38646 protein (i.e., "the coding region," from nucleotides 115-2413 of SEQ ID NO: 1), as well as 5'-non-translated sequences (nucleotides 1-114 of SEQ ID NO: 1) or 3'-non-translated sequences (nucleotides 2414-2561 of SEQ ID NO: 1). Alternatively, the nucleic acid molecule can include only the coding region of SEQ ID NO: 1 (e.g., nucleotides 115-2413, corresponding to SEQ ID NO: 3) and, e.g., no flanking sequences which normally accompany the subject

sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to the 766 amino acid residue protein of SEQ ID NO: 2.

In another embodiment, an isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in one of SEQ ID NO: 1, SEQ ID NO: 3, and a portion of either of those sequences. In other embodiments, the nucleic acid molecule of the invention is sufficiently complementary to the nucleotide sequence shown in one of SEQ ID NO: 1 and SEQ ID NO: 3 that it can hybridize with a nucleic acid having that sequence, thereby forming a stable duplex.

In one embodiment, an isolated nucleic acid molecule of the invention includes a nucleotide sequence which is at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more homologous to the entire length of the nucleotide sequence shown in one of SEQ ID NO: 1, SEQ ID NO: 3, and a portion, preferably of the same length, of either of these nucleotide sequences.

38646 Nucleic Acid Fragments

A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of one of SEQ ID NOs: 1 and 3. For example, such a nucleic acid molecule can include a fragment that can be used as a probe or primer or a fragment encoding a portion of a 38646 protein, e.g., an immunogenic or biologically active portion of a 38646 protein. A fragment can comprise nucleotide residues corresponding to residues 210-392 of SEQ ID NO: 2, which encodes a RhoGEF domain of human 38646, nucleotide residues corresponding to residues 423-521 of SEQ ID NO: 2, which encodes a PH domain of human 38646, nucleotide residues corresponding to residues 644-740 of SEQ ID NO: 2, which encodes another PH domain of human 38646, or nucleotide residues corresponding to residues 554-620 of SEQ ID NO: 2, which encodes an FYVE domain of human 38646. The nucleotide sequence determined from the cloning of the 38646 gene facilitates generation of probes and primers for use in identifying and/or cloning other 38646 family members, or fragments thereof, as well as 38646 homologues, or fragments thereof, from other species.

In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5'- or 3'-non-

coding region. Other embodiments include a fragment that includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof that are at least about 250 amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or more domain, region, or functional site described herein.

38646 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of one of SEQ ID NO: 1, SEQ ID NO: 3, and a naturally occurring allelic variant or mutant of SEQ ID NO: 1 or SEQ ID NO: 3.

In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length. It should be identical, or differ by 1, or fewer than 5 or 10 bases, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid that encodes: a RhoGEF domain at about amino acid residues 210-392 of SEQ ID NO: 2; a PH domain at about amino acid residues 423 to 521, or 644 to 740 of SEQ ID NO: 2; or an FYVE domain at about amino acid residues 554 to 620 of SEQ ID NO: 2.

In another embodiment a set of primers is provided, e.g., primers suitable for use in a PCR, which can be used to amplify a selected region of a 38646 sequence. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differs by one base from a sequence disclosed

herein or from a naturally occurring variant. Primers suitable for amplifying all or a portion of any of the following regions are provided: e.g., one or more a RhoGEF domain, a PH domain, and an FYVE domain, all as defined above relative to SEQ ID NO: 2.

A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

A nucleic acid fragment encoding a "biologically active portion of a 38646 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of one of SEQ ID NO: 1 and SEQ ID NO: 3, which encodes a polypeptide having a 38646 biological activity (e.g., the biological activities of the 38646 proteins that are described herein, such as actin filament binding activity or ability to induce cytoskeletal organization and intracellular protein and vesicle localization), expressing the encoded portion of the 38646 protein (e.g., by recombinant expression in vitro) and assessing the activity of the encoded portion of the 38646 protein. For example, a nucleic acid fragment encoding a biologically active portion of 38646 includes a RhoGEF domain, e.g., amino acid residues 210 to 392 of SEQ ID NO: 2, a PH domain, e.g., amino acid residues 423 to 521, or 644 to 740 of SEQ ID NO: 2; or an FYVE domain, e.g., amino acid residues 554 to 620 of SEQ ID NO: 2. A nucleic acid fragment encoding a biologically active portion of a 38646 polypeptide can comprise a nucleotide sequence that is greater than 25 or more nucleotides in length.

In one embodiment, a nucleic acid includes one that has a nucleotide sequence which is greater than 260, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 2000, or 2500 or more nucleotides in length and that hybridizes under stringent hybridization conditions with a nucleic acid molecule having the sequence of one of SEQ ID NO: 1 and SEQ ID NO: 3. Preferably the 38646 nucleic acid has a nucleotide sequence that includes at least 260, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 2000, or 2500 or more consecutive nucleotide residues of one of SEQ ID NOs: 1 and 3.

38646 Nucleic Acid Variants

The invention further encompasses nucleic acid molecules having a sequence that differs from the nucleotide sequence shown in one of SEQ ID NO: 1 and SEQ ID NO: 3. Such differences can be attributable to degeneracy of the genetic code (i.e., differences which

result in a nucleic acid that encodes the same 38646 proteins as those encoded by the nucleotide sequence disclosed herein). In another embodiment, an isolated nucleic acid molecule of the invention encodes a protein having an amino acid sequence which differs by at least 1, but by fewer than 5, 10, 20, 50, or 100, amino acid residues from SEQ ID NO: 2. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

Nucleic acids of the inventor can be chosen for having codons, which are preferred, or non-preferred, for a particular expression system. For example, the nucleic acid can be one in which at least one colon, at preferably at least 10%, or 20% of the codons has been altered such that the sequence is optimized for expression in E. coli, yeast, human, insect, or CHO cells.

Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

In a preferred embodiment, the nucleic acid has a sequence that differs from that of one of SEQ ID NO: 1 and SEQ ID NO: 3, e.g., as follows: by at least one, but by fewer than 10, 20, 30, or 40, nucleotide residues; or by at least one but by fewer than 1%, 5%, 10% or 20% of the nucleotide residues in the subject nucleic acid. If necessary for this analysis the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more identical to the nucleotide sequence shown in one of SEQ ID NO: 1, SEQ ID NO: 3, or a fragment of one of these sequences. Such nucleic acid

molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in one of SEQ ID NO: 1, SEQ ID NO: 3, or a fragment of one of these sequences. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the 38646 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the 38646 gene.

Preferred variants include those that are correlated with any of the 38646 biological activities described herein, e.g., catalyzing activation of a Rho GTP-binding protein.

Allelic variants of 38646 (e.g., human 38646) include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the 38646 protein within a population that maintain the ability to mediate any of the 38646 biological activities described herein.

Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO: 2, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the 38646 (e.g., human 38646) protein within a population that do not have the ability to mediate any of the 38646 biological activities described herein. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO: 2, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

Moreover, nucleic acid molecules encoding other 38646 family members and, thus, which have a nucleotide sequence which differs from the 38646 sequences of one of SEQ ID NO: 1 and SEQ ID NO: 3 are within the scope of the invention.

Antisense Nucleic Acid Molecules, Ribozymes and Modified 38646 Nucleic Acid Molecules

In another aspect, the invention features, an isolated nucleic acid molecule that is antisense to 38646. An "antisense" nucleic acid can include a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The

antisense nucleic acid can be complementary to an entire 38646 coding strand, or to only a portion thereof (e.g., the coding region of human 38646 corresponding to SEQ ID NO: 3). In another embodiment, the antisense nucleic acid molecule is antisense to a "non-coding region" of the coding strand of a nucleotide sequence encoding 38646 (e.g., the 5'- and 3'-non-translated regions).

An antisense nucleic acid can be designed such that it is complementary to the entire coding region of 38646 mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or non-coding region of 38646 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 38646 mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, or 80 or more nucleotide residues in length.

An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been sub-cloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject (e.g., by direct injection at a tissue site), or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 38646 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface,

e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an alpha-anomeric nucleic acid molecule. An alpha-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual beta-units, the strands run parallel to each other (Gaultier et al., 1987, Nucl. Acids. Res. 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res. 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a 38646-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequence of a 38646 cDNA disclosed herein (i.e., SEQ ID NO: 1 or SEQ ID NO: 3), and a sequence having known catalytic sequence responsible for mRNA cleavage (see, for example, U.S. Patent number 5,093,246 or Haselhoff et al. (1988, Nature 334:585-591). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 38646-encoding mRNA (e.g., U.S. Patent number 4,987,071; and U.S. Patent number 5,116,742). Alternatively, 38646 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (e.g., Bartel et al., 1993, Science 261:1411-1418).

38646 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 38646 (e.g., the 38646 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 38646 gene in target cells (Helene, 1991, Anticancer Drug Des. 6:569-584; Helene, et al., 1992, Ann. N.Y. Acad. Sci. 660:27-36; Maher, 1992, Bioassays 14:807-815). The potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5' to 3', 3' to 5' manner,

such that they hybridize with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

A 38646 nucleic acid molecule can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (Hyrup et al., 1996, Bioorg. Med. Chem. 4:5-23). As used herein, the terms "peptide nucleic acid" (PNA) refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996, supra; Perry-O'Keefe et al., Proc. Natl. Acad. Sci. USA 93:14670-14675).

PNAs of 38646 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or anti-gene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 38646 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases, as described in Hyrup et al., 1996, supra); or as probes or primers for DNA sequencing or hybridization (Hyrup et al., 1996, supra; Perry-O'Keefe, supra).

In other embodiments, the oligonucleotide can include other appended groups such as peptides (e.g., for targeting host cell receptors in vivo), or agents facilitating transport across the cell membrane (e.g., Letsinger et al., 1989, Proc. Natl. Acad. Sci. USA 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. USA 84:648-652; PCT publication number WO 88/09810) or the blood-brain barrier (see, e.g., PCT publication number WO 89/10134).

In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (e.g., Krol et al., 1988, Bio-Techniques 6:958-976) or intercalating agents (e.g., Zon, 1988, Pharm. Res. 5:539-549). To this end, the oligonucleotide can be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 38646 nucleic acid of the invention, two complementary regions, one having a fluorophore and the other having a quencher, such that the molecular beacon is useful for quantitating the presence of the 38646 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in U.S. Patent number. 5,854,033, U.S. Patent number 5,866,336, and U.S. Patent number 5,876,930.

Isolated 38646 Polypeptides

In another aspect, the invention features, an isolated 38646 protein, or fragment, e.g., a biologically active portion, for use as immunogens or antigens to raise or test (or more generally to bind) anti-38646 antibodies. 38646 protein can be isolated from cells or tissue sources using standard protein purification techniques. 38646 protein or fragments thereof can be produced by recombinant DNA techniques or synthesized chemically.

Polypeptides of the invention include those that arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and post-translational events. The polypeptide can be expressed in systems, e.g., cultured cells, which result in substantially the same post-translational modifications present when the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of post-translational modifications, e.g., glycosylation or cleavage, present when expressed in a native cell.

In a preferred embodiment, a 38646 polypeptide has one or more of the following characteristics:

(1) it catalyzes activation of Rho GTP-binding proteins, e.g., by displacing GDP bound with the inactive form of the Rho protein with GTP;

(2) it modulates cell shape;

(3) it modulates cell growth;

(4) it modulates cell deformation;

(5) it modulates cell differentiation;

(6) it modulates cell adhesion;

(7) it modulates infiltration of a cell into or through a tissue;

(8) it modulates progression of a cell through the cell cycle;

(9) it modulates embryonic development

(10) it modulates angiogenesis;

(11) it modulates neurite growth;

(12) it modulates tumor cell invasion or metastasis;

(13) it modulates tumorigenesis;

(14) it modulates wound healing;

(15) it modulates bone remodeling;

(16) it modulates apoptosis;

(17) it modulates the tensile strength of a tissue;

(18) it modulates the binding affinity of a cell for another cell or an extracellular matrix;

(19) it has a molecular weight, amino acid composition or other physical characteristic of a 38646 protein of SEQ ID NO: 2;

(20) it has an overall sequence similarity (identity) of at least 60-65%, preferably at least 70%, more preferably at least 75, 80, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% or more, with a portion of SEQ ID NO: 2;

(21) it has a RhoGEF domain which is preferably about 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or higher, identical with amino acid residues 210-392 of SEQ ID NO: 2;

(22) it has a PH domain which is preferably about 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or higher, identical with amino acid residues 423-521 of SEQ ID NO: 2;

(23) it has a PH domain which is preferably about 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or higher, identical with amino acid residues 644-740 of SEQ ID NO: 2; and

(24) it has a RhoGEF domain which is preferably about 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or higher, identical with amino acid residues 554-620 of SEQ ID NO: 2.

In a preferred embodiment, the 38646 protein or fragment thereof differs only insubstantially, if at all, from the corresponding sequence in SEQ ID NO: 2. In one embodiment, it differs by at least one, but by fewer than 15, 10 or 5 amino acid residues. In another, it differs from the corresponding sequence in SEQ ID NO: 2 by at least one residue but fewer than 20%, 15%, 10% or 5% of the residues differ from the corresponding sequence in SEQ ID NO: 2 (if this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences). The differences are, preferably, differences or changes at a non-essential amino acid residues or involve a conservative substitution of one residue for another. In a preferred embodiment the differences are not in residues 39 to 225 of SEQ ID NO: 2.

Other embodiments include a protein that has one or more changes in amino acid sequence, relative to SEQ ID NO: 2 (e.g., a change in an amino acid residue which is not essential for activity). Such 38646 proteins differ in amino acid sequence from SEQ ID NO: 2, yet retain biological activity.

In one embodiment, the protein includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to SEQ ID NO: 2.

A 38646 protein or fragment is provided which has an amino acid sequence which varies from SEQ ID NO: 2 in one or both of the regions corresponding to residues 1-38 and 226-520 of SEQ ID NO: 2 by at least one, but by fewer than 15, 10 or 5 amino acid residues, but which does not differ from SEQ ID NO: 2 in the region corresponding to residues 39-225 of SEQ ID NO: 2 (if this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences). In some embodiments the difference is at a non-

essential residue or is a conservative substitution, while in others the difference is at an essential residue or is a non-conservative substitution.

A biologically active portion of a 38646 protein should include at least the 38646 RhoGEF domain, and preferably also includes the PH domains and the FYVE domain of naturally-occurring 38646 protein. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native 38646 protein.

In a preferred embodiment, the 38646 protein has the amino acid sequence SEQ ID NO: 2. In other embodiments, the 38646 protein is substantially identical to SEQ ID NO: 2. In yet another embodiment, the 38646 protein is substantially identical to SEQ ID NO: 2 and retains the functional activity of the protein of SEQ ID NO: 2.

38646 Chimeric or Fusion Proteins

In another aspect, the invention provides 38646 chimeric or fusion proteins. As used herein, a 38646 "chimeric protein" or "fusion protein" includes a 38646 polypeptide linked to a non-38646 polypeptide. A "non-38646 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the 38646 protein, e.g., a protein which is different from the 38646 protein and which is derived from the same or a different organism. The 38646 polypeptide of the fusion protein can correspond to all or a portion e.g., a fragment described herein of a 38646 amino acid sequence. In a preferred embodiment, a 38646 fusion protein includes at least one or more biologically active portions of a 38646 protein. The non-38646 polypeptide can be fused to the amino or carboxyl terminus of the 38646 polypeptide.

The fusion protein can include a moiety that has a high affinity for a ligand. For example, the fusion protein can be a GST-38646 fusion protein in which the 38646 sequences are fused to the carboxyl terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 38646. Alternatively, the fusion protein can be a 38646 protein containing a heterologous signal sequence at its amino terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 38646 can be increased through use of a heterologous signal sequence.

Fusion proteins can include all or a part of a serum protein, e.g., an IgG constant region, or human serum albumin.

The 38646 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject in vivo. The 38646 fusion proteins can be used to affect the bioavailability of a 38646 substrate. 38646 fusion proteins can be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 38646 protein; (ii) mis-regulation of the 38646 gene; and (iii) aberrant post-translational modification of a 38646 protein.

Moreover, the 38646-fusion proteins of the invention can be used as immunogens to produce anti-38646 antibodies in a subject, to purify 38646 ligands and in screening assays to identify molecules that inhibit the interaction of 38646 with a 38646 substrate.

Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 38646-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 38646 protein.

Variants of 38646 Proteins

In another aspect, the invention also features a variant of a 38646 polypeptide, e.g., which functions as an agonist (mimetics) or as an antagonist. Variants of the 38646 proteins can be generated by mutagenesis, e.g., discrete point mutation, the insertion or deletion of sequences or the truncation of a 38646 protein. An agonist of the 38646 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 38646 protein. An antagonist of a 38646 protein can inhibit one or more of the activities of the naturally occurring form of the 38646 protein by, for example, competitively modulating a 38646-mediated activity of a 38646 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 38646 protein.

Variants of a 38646 protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 38646 protein for agonist or antagonist activity.

Libraries of fragments e.g., amino-terminal, carboxyl-terminal, or internal fragments, of a 38646 protein coding sequence can be used to generate a variegated population of fragments for screening and subsequent selection of variants of a 38646 protein.

Variants in which a cysteine residue is added or deleted or in which a residue that is glycosylated is added or deleted are particularly preferred.

Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 38646 variants (Arkin et al., 1992, Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al., 1993, Protein Engr. 6:327-331).

Cell based assays can be exploited to analyze a variegated 38646 library. For example, a library of expression vectors can be transfected into a cell line, e.g., a cell line, which ordinarily responds to 38646 in a substrate-dependent manner. The transfected cells are then contacted with 38646 and the effect of the expression of the mutant on signaling by the 38646 substrate can be detected, e.g., by measuring changes in cell growth and/or enzymatic activity. Plasmid DNA can then be recovered from the cells that score for inhibition, or alternatively, potentiation of signaling by the 38646 substrate, and the individual clones further characterized.

In another aspect, the invention features a method of making a 38646 polypeptide, e.g., a peptide having a non-wild-type activity, e.g., an antagonist, agonist, or super agonist of a naturally-occurring 38646 polypeptide, e.g., a naturally-occurring 38646 polypeptide. The method includes: altering the sequence of a 38646 polypeptide, e.g., altering the sequence, e.g., by substitution or deletion of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

In another aspect, the invention features a method of making a fragment or analog of a 38646 polypeptide a biological activity of a naturally occurring 38646 polypeptide. The method includes: altering the sequence, e.g., by substitution or deletion of one or more residues, of a 38646 polypeptide, e.g., altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

Anti-38646 Antibodies

In another aspect, the invention provides an anti-38646 antibody. The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin.

The antibody can be a polyclonal, monoclonal, recombinant, e.g., a chimeric or humanized, fully-human, non-human, e.g., murine, or single chain antibody. In a preferred embodiment, it has effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

A full-length 38646 protein or, antigenic peptide fragment of 38646 can be used as an immunogen or can be used to identify anti-38646 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptide of 38646 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO: 2 and encompasses an epitope of 38646. Preferably, the antigenic peptide includes at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Fragments of 38646 which include about residues 39-225 of SEQ ID NO: 2 can be used to make antibodies, e.g., for use as immunogens or to characterize the specificity of an antibody, against hydrophobic regions of the 38646 protein. Similarly, a fragment of 38646 which include about residues 210-230 or 300-320 of SEQ ID NO: 2 can be used to make an antibody against a hydrophilic region of the 38646 protein.

Antibodies reactive with, or specific for, any of these regions, or other regions or domains described herein are provided.

Preferred epitopes encompassed by the antigenic peptide are regions of 38646 are located on the surface of the protein, e.g., hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 38646 protein sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the 38646 protein and are thus likely to constitute surface residues useful for targeting antibody production.

In a preferred embodiment the antibody binds an epitope on any domain or region on 38646 proteins described herein.

Chimeric, humanized, but most preferably, completely human antibodies are desirable for applications which include repeated administration, e.g., therapeutic treatment (and some diagnostic applications) of human patients.

The anti-38646 antibody can be a single chain antibody. A single-chain antibody (scFV) can be engineered (e.g., Colcher et al., 1999, Ann. N.Y. Acad. Sci. 880:263-280; Reiter, 1996, Clin. Cancer Res. 2:245-252). The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target 38646 protein.

In a preferred embodiment, the antibody has reduced or no ability to bind an Fc receptor. For example, it can be an isotype, subtype, fragment or other mutant, which does not support binding to an Fc receptor, e.g., it can have a mutated or deleted Fc receptor binding region.

An anti-38646 antibody (e.g., monoclonal antibody) can be used to isolate 38646 by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-38646 antibody can be used to detect 38646 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-38646 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include

horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses.

A vector can include a 38646 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (e.g., 38646 proteins, mutant forms of 38646 proteins, fusion proteins, and the like).

The recombinant expression vectors of the invention can be designed for expression of 38646 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in E. coli, insect cells (e.g., using baculovirus expression

vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel (1990, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in E. coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith et al., 1988, Gene 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be used in 38646 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 38646 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone marrow cells that are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six weeks).

To maximize recombinant protein expression in E. coli, the protein is expressed in a host bacterial strain with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, 1990, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in E. coli (Wada et al., 1992, Nucl. Acids Res.

20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

The 38646 expression vector can be a yeast expression vector, a vector for expression in insect cells, e.g., a baculovirus expression vector, or a vector suitable for expression in mammalian cells.

When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used viral promoters are derived from polyoma, adenovirus 2, cytomegalovirus and simian virus 40 (SV40).

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al., 1987, Genes Dev. 1:268-277), lymphoid-specific promoters (Calame et al., 1988, Adv. Immunol. 43:235-275), in particular promoters of T cell receptors (Winoto et al., 1989, EMBO J. 8:729-733) and immunoglobulins (Banerji et al., 1983, Cell 33:729-740; Queen et al., 1983, Cell 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne et al., 1989, Proc. Natl. Acad. Sci. USA 86:5473-5477), pancreas-specific promoters (Edlund et al., 1985, Science 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent number 4,873,316 and European Patent Application publication number 264,166). Developmentally-regulated promoters are also encompassed, for example, the murine hox promoters (Kessel et al., 1990, Science 249:374-379) and the alpha-fetoprotein promoter (Campes et al., 1989, Genes Dev. 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes, see Weintraub, H. et al. (1986, Trends Genet. 1:Review).

Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein, e.g., a 38646 nucleic acid molecule within a recombinant expression vector or a 38646 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell, but also to the progeny or potential progeny of such a cell. Because certain modifications can occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a 38646 protein can be expressed in bacterial cells such as E. coli, insect cells, yeast or mammalian cells (such as Chinese hamster ovary (CHO) cells) or COS cells. Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

A host cell of the invention can be used to produce (i.e., express) a 38646 protein. Accordingly, the invention further provides methods for producing a 38646 protein using the host cells of the invention. In one embodiment, the method includes culturing the host cell of the invention (into which a recombinant expression vector encoding a 38646 protein has been introduced) in a suitable medium such that a 38646 protein is produced. In another embodiment, the method further includes isolating a 38646 protein from the medium or the host cell.

In another aspect, the invention features, a cell or purified preparation of cells which include a 38646 transgene, or which otherwise mal-express 38646. The cell preparation can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a 38646 transgene, e.g., a heterologous form of a 38646, e.g., a gene derived from humans (in the case of a non-human

cell). The 38646 transgene can be mal-expressed, e.g., over-expressed or under-expressed. In other preferred embodiments, the cell or cells include a gene that mal-expresses an endogenous 38646, e.g., a gene the expression of which is disrupted, e.g., a knockout. Such cells can serve as a model for studying disorders that are related to mutated or mal-expressed 38646 alleles or for use in drug screening.

In another aspect, the invention includes, a human cell, e.g., a hematopoietic stem cell, transformed with nucleic acid that encodes a subject 38646 polypeptide.

Also provided are cells, preferably human cells, e.g., human hematopoietic or fibroblast cells, in which an endogenous 38646 is under the control of a regulatory sequence that does not normally control expression of the endogenous 38646 gene. The expression characteristics of an endogenous gene within a cell, e.g., a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous 38646 gene. For example, an endogenous 38646 gene that is "transcriptionally silent," e.g., not normally expressed, or expressed only at very low levels, can be activated by inserting a regulatory element that is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombination, can be used to insert the heterologous DNA as described (e.g., U.S. Patent number 5,272,071; PCT publication number WO 91/06667).

Transgenic Animals

The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of a 38646 protein and for identifying and/or evaluating modulators of 38646 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell

types or tissues of the transgenic animal, other transgenes, e.g., a knockout, reduce expression. Thus, a transgenic animal can be one in which an endogenous 38646 gene has been altered, e.g., by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal (e.g., an embryonic cell of the animal, prior to development of the animal).

Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of a 38646 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 38646 transgene in its genome and/or expression of 38646 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 38646 protein can further be bred to other transgenic animals carrying other transgenes.

38646 proteins or polypeptides can be expressed in transgenic animals or plants, e.g., a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk- or egg-specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

The invention also includes a population of cells from a transgenic animal, as discussed, e.g., below.

Uses

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic). The isolated nucleic acid molecules of the invention can be used, for example, to express a 38646 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect a 38646 mRNA (e.g., in a biological sample), to detect a genetic alteration in a 38646 gene and to modulate 38646 activity, as described further below. The 38646 proteins can be

used to treat disorders characterized by insufficient or excessive production of a 38646 substrate or production of 38646 inhibitors. In addition, the 38646 proteins can be used to screen for naturally occurring 38646 substrates, to screen for drugs or compounds which modulate 38646 activity, as well as to treat disorders characterized by insufficient or excessive production of 38646 protein or production of 38646 protein forms which have decreased, aberrant or unwanted activity compared to 38646 wild-type protein. Exemplary disorders include those described herein). Moreover, the anti-38646 antibodies of the invention can be used to detect and isolate 38646 proteins, regulate the bioavailability of 38646 proteins, and modulate 38646 activity.

A method of evaluating a compound for the ability to interact with, e.g., bind to, a subject 38646 polypeptide is provided. The method includes: contacting the compound with the subject 38646 polypeptide; and evaluating the ability of the compound to interact with, e.g., to bind or form a complex with, the subject 38646 polypeptide. This method can be performed in vitro, e.g., in a cell free system, or in vivo, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally-occurring molecules that interact with a subject 38646 polypeptide. It can also be used to find natural or synthetic inhibitors of a subject 38646 polypeptide. Screening methods are discussed in more detail below.

Screening Assays

The invention provides screening methods (also referred to herein as "assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind with 38646 proteins, have a stimulatory or inhibitory effect on, for example, 38646 expression or 38646 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 38646 substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., 38646 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

In one embodiment, the invention provides assays for screening candidate or test compounds that are substrates of a 38646 protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test

compounds that bind to or modulate the activity of a 38646 protein or polypeptide or a biologically active portion thereof.

In another embodiment, the invention provides assays for screening candidate or test compounds that mimic a substrate of a 38646 protein or which otherwise interfere with the normal interaction between 38646 protein and its physiological substrate. 38646 protein can interact with and reorganize cytoskeleton components such as actin, thereby facilitating cell shape alterations (e.g., dendrite formation, filipodia formation, and other cellular extensions), increasing cellular motility, and altering intracellular localization of proteins and secretory vesicles. Candidate or test compounds which mimic or duplicate part of the chemical structure of an actin molecule or filament can therefore be screened to assess their effect on the actin filament binding or cytoskeleton organization activity of 38646 protein. Preferably, such candidate or test compounds mimic or duplicate a portion of a cytoskeleton component that normally comprises a GEF recognition site, except that in the candidate or test compound this site is altered. These screening methods can be used to assess the effectiveness of a candidate or test compound for modulating cell shape, motility, intracellular organization, and secretion release rate. Thus, for example, these methods are suitable both for assessing the ability of a test or candidate compound to enhance the ability of neurons to extend to or toward other neurons and the rate of synaptic vesicle release and for assessing the ability of a test or candidate compound to inhibit or alter the secretions of a gland (e.g., pancreas or liver).

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive; e.g., Zuckermann et al., 1994, J. Med. Chem. 37:2678-2685); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, 1997, Anticancer Drug Des. 12:145).

Examples of methods for the synthesis of molecular libraries have been described (e.g., DeWitt et al., 1993, Proc. Natl. Acad. Sci. USA 90:6909; Erb et al., 1994, Proc. Natl. Acad. Sci. USA 91:11422; Zuckermann et al., 1994, J. Med. Chem. 37:2678; Cho et al., 1993, Science 261:1303; Carrell et al., 1994, Angew. Chem. Int. Ed. Engl. 33:2059; Carell et al., 1994, Angew. Chem. Int. Ed. Engl. 33:2061; and Gallop et al., 1994, J. Med. Chem. 37:1233).

Libraries of compounds can be presented in solution (e.g., Houghten, 1992, Biotechniques 13:412-421), or on beads (Lam, 1991, Nature 354:82-84), chips (Fodor, 1993, Nature 364:555-556), bacteria (U.S. Patent number 5,223,409), spores (U.S. Patent number 5,223,409), plasmids (Cull et al., 1992, Proc. Natl. Acad. Sci. USA 89:1865-1869), or on phage (Scott et al., 1990, Science 249:386-390; Devlin, 1990, Science 249:404-406; Cwirla et al., 1990, Proc. Natl. Acad. Sci. USA 87:6378-6382; Felici, 1991, J. Mol. Biol. 222:301-310; U.S. Patent number 5,223,409).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a 38646 protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate 38646 activity is determined. Determining the ability of the test compound to modulate 38646 activity can be accomplished by monitoring, for example, changes in enzymatic activity. The cell, for example, can be of mammalian origin.

The ability of the test compound to modulate 38646 binding to a compound, e.g., a 38646 substrate, or to bind to 38646 can also be evaluated. This can be accomplished, for example, by coupling the compound, e.g., the substrate, with a radioisotope or enzymatic label such that binding of the compound, e.g., the substrate, to 38646 can be determined by detecting the labeled compound, e.g., substrate, in a complex. Alternatively, 38646 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate 38646 binding to a 38646 substrate in a complex. For example, compounds (e.g., 38646 substrates) can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radio-emission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish

peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

The ability of a compound (e.g., a 38646 substrate) to interact with 38646 with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with 38646 without the labeling of either the compound or the 38646 (McConnell et al., 1992, Science 257:1906-1912). As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and 38646.

In yet another embodiment, a cell-free assay is provided in which a 38646 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 38646 protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the 38646 proteins to be used in assays of the present invention include fragments that participate in interactions with non-38646 molecules, e.g., fragments with high surface probability scores.

Soluble and/or membrane-bound forms of isolated proteins (e.g., 38646 proteins or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it can be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)n, 3-{(3-cholamidopropyl) dimethylamminio}-1-propane sulfonate (CHAPS), 3-{(3-cholamidopropyl) dimethylamminio}-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET; e.g., U.S. Patent number 5,631,169; U.S. Patent number

4,868,103). A fluorophore label is selected such that a first donor molecule's emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule can simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label can be differentiated from that of the 'donor.' Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

In another embodiment, determining the ability of the 38646 protein to bind to a target molecule can be accomplished using real-time biomolecular interaction analysis (BIA; e.g., Sjolander et al., 1991, Anal. Chem. 63:2338-2345; Szabo et al., 1995, Curr. Opin. Struct. Biol. 5:699-705). "Surface plasmon resonance" (SPR) or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of SPR), resulting in a detectable signal that can be used as an indication of real-time reactions between biological molecules.

In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

It can be desirable to immobilize either 38646, an anti-38646 antibody or its target molecule to facilitate separation of complexed from non-complexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 38646 protein, or interaction of a 38646 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and

micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/38646 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione Sepharose™ beads (Sigma Chemical, St. Louis, MO) or glutathione-derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 38646 protein, and the mixture incubated under conditions conducive for complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 38646 binding or activity determined using standard techniques.

Other techniques for immobilizing either a 38646 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated 38646 protein or target molecules can be prepared from biotin- N-hydroxy-succinimide using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96-well plates (Pierce Chemical).

In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, non-reacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

In one embodiment, this assay is performed utilizing antibodies reactive with 38646 protein or target molecules but which do not interfere with binding of the 38646 protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound

target or 38646 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 38646 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 38646 protein or target molecule.

Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from non-reacted components, by any of a number of standard techniques, including, but not limited to: differential centrifugation (e.g., Rivas et al., 1993, Trends Biochem. Sci. 18:284-287); chromatography (e.g., gel filtration chromatography or ion-exchange chromatography); electrophoresis (e.g., Ausubel et al., eds., 1999, Current Protocols in Molecular Biology, J. Wiley, New York); and immunoprecipitation (e.g., Ausubel, supra). Such resins and chromatographic techniques are known to one skilled in the art (e.g., Heegaard, 1998, J. Mol. Recognit. 11:141-148; Hage et al., 1997, J. Chromatogr. B Biomed. Sci. Appl. 699:499-525). Further, fluorescence energy transfer can also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

In a preferred embodiment, the assay includes contacting the 38646 protein or biologically active portion thereof with a known compound which binds 38646 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 38646 protein, wherein determining the ability of the test compound to interact with a 38646 protein includes determining the ability of the test compound to preferentially bind to 38646 or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

The target gene products of the invention can, in vivo, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the 38646 genes herein identified. In an alternative embodiment, the invention

provides methods for determining the ability of the test compound to modulate the activity of a 38646 protein through modulation of the activity of a downstream effector of a 38646 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test

compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, non-reacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from non-reacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a pre-formed complex of the target gene product and the interactive cellular or

extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (e.g., U.S. Patent number 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

In yet another aspect, the 38646 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (e.g., U.S. Patent number 5,283,317; Zervos et al., 1993, Cell 72:223-232; Madura et al., 1993, J. Biol. Chem. 268:12046-12054; Bartel et al., 1993, Biotechniques 14:920-924; Iwabuchi et al., 1993, Oncogene 8:1693-1696; PCT publication number WO 94/10300), to identify other proteins, which bind to or interact with 38646 ("38646-binding proteins" or "38646-bp") and are involved in 38646 activity. Such 38646-bps can be activators or inhibitors of signals by the 38646 proteins or 38646 targets as, for example, downstream elements of a 38646-mediated signaling pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 38646 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively, the 38646 protein can be fused to the activator domain). If the "bait" and the "prey" proteins are able to interact in vivo forming a 38646-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein that interacts with the 38646 protein.

In another embodiment, modulators of 38646 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression

of 38646 mRNA or protein evaluated relative to the level of expression of 38646 mRNA or protein in the absence of the candidate compound. When expression of 38646 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 38646 mRNA or protein expression. Alternatively, when expression of 38646 mRNA or protein is less (i.e., statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 38646 mRNA or protein expression. The level of 38646 mRNA or protein expression can be determined by methods described herein for detecting 38646 mRNA or protein.

In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a 38646 protein can be confirmed in vivo, e.g., in an animal such as an animal model for a disease.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a 38646 modulating agent, an antisense 38646 nucleic acid molecule, a 38646-specific antibody, or a 38646-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used for treatments as described herein.

Detection Assays

Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome, e.g., to locate gene regions associated with genetic disease or to associate 38646 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

Chromosome Mapping

The 38646 nucleotide sequences or portions thereof can be used to map the location of the 38646 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 38646 sequences with genes associated with disease.

Briefly, 38646 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 base pairs in length) from the 38646 nucleotide sequence (e.g., SEQ ID NO: 1 or SEQ ID NO: 3). These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 38646 sequences will yield an amplified fragment.

A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes (D'Eustachio et al., 1983, Science 220:919-924).

Other mapping strategies e.g., in situ hybridization as described (Fan et al., 1990, Proc. Natl. Acad. Sci. USA 87:6223-6227), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map 38646 to a chromosomal location.

Fluorescence in situ hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of FISH, see Verma et al. (1988, Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to non-coding regions of the genes are typically preferred for mapping purposes. Coding sequences are more likely to be

conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data (such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), as described (e.g., Egeland et al., 1987, Nature, 325:783-787).

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 38646 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

38646 sequences can be used to identify individuals from biological samples using, e.g., restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, e.g., in a Southern blot, and probed to yield bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent number 5,272,057).

Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 38646 nucleotide sequence described herein can be used to prepare PCR primers homologous to the 5'- and 3'-ends of the sequence. These primers can then be used to

amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the non-coding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the non-coding regions, fewer sequences are necessary to differentiate individuals. The non-coding sequences of SEQ ID NO: 1 can provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a non-coding amplified sequence of 100 bases. If predicted coding sequences are used, such as those in SEQ ID NO: 3, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from 38646 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

Use of Partial 38646 Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e., another DNA sequence that is unique to a particular individual). As mentioned above, actual nucleotide sequence information can be used for identification as

an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to non-coding regions of SEQ ID NO: 1 (e.g., fragments having a length of at least 20 nucleotide residues, preferably at least 30 nucleotide residues) are particularly appropriate for this use.

The 38646 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or label-able probes which can be used in, for example, an in situ hybridization technique, to identify a specific tissue, e.g., a tissue containing hematopoietic cells. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 38646 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., 38646 primers or probes can be used to screen tissue culture for contamination (i.e., to screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

Generally, the invention provides a method of determining if a subject is at risk for a disorder related to a lesion in, or the malexpression of, a gene that encodes a 38646 polypeptide.

Such disorders include, e.g., a disorder associated with the malexpression of a 38646 polypeptide, e.g., a developmental disorder, a neuronal disorder, a glandular disorder, an immune disorder or a neoplastic disorder.

As the data disclosed herein indicate, expression of the 38646 gene is associated with the ability of nerve tissue to alter cell shape, cell motility, cytoskeleton organization, and synaptic vesicle localization. Thus, detection (organism-wide or in a neuronal tissue or cell type) of mutations which affect expression of the 38646 gene can indicate whether a subject is likely to develop a neuronal disorder. Similarly, detection of expression levels of 38646 protein in neuronal tissues can indicate the degree or extent of damage that has resulted from a

traumatic injury to the brain or spine or from an ischemic injury such (e.g., resulting from stroke) and is therefore useful as a prognostic indicator for recovery.

The data disclosed herein also indicate a role for 38646 molecules in maintaining the normal function of vascular tissues (e.g., maintaining normal binding between arterial endothelial cells and maintaining the ability of such cells to separate in the presence of an extravasating leukocyte). Thus, detection (organism-wide or in a vascular tissue or cell type) of mutations which affect expression of the 38646 gene can indicate whether a subject is likely to develop a vascular disorder.

The data disclosed herein indicate that 38646 modulates binding between cells and between a cell and its extracellular matrix. Such modulation includes modulation of the tensile strength and structural integrity of a solid tissue. 38646 molecules can be used to modulate 38646 expression and activity, and can be used to alleviate, inhibit, or prevent disorders that are characterized by aberrant strength, integrity, or permeability of a tissue. Examples of such disorders include a variety of kidney disorders (e.g., chronic and acute renal failure, urinary tract infections, immunologically-mediated renal diseases, glomerular diseases such as acute nephritic syndrome, membranoproliferative glomerulonephritis type I, nephrotic syndrome, and tubulointerstitial disorders such as chronic tubulointerstitial nephritis). Other examples include tumor metastasis, growth, and spread.

The method includes one or more of the following:

(i) detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 38646 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, e.g., a mutation in the 5'-control region;

(ii) detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 38646 gene;

(iii) detecting, in a tissue of the subject, the malexpression of the 38646 gene at the mRNA level, e.g., detecting a non-wild-type level of a mRNA; and

(iv) detecting, in a tissue of the subject, the malexpression of the gene at the protein level, e.g., detecting a non-wild-type level of a 38646 polypeptide.

In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 38646 gene; an insertion of one or

more nucleotides into the gene, a point mutation, e.g., a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO: 1, or naturally occurring mutants thereof, or 5'- or 3'-flanking sequences naturally associated with the 38646 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting the presence or absence of the genetic lesion by hybridization of the probe/primer to the nucleic acid, e.g., by in situ hybridization.

In preferred embodiments, detecting the malexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 38646 gene; the presence of a non-wild-type splicing pattern of a messenger RNA transcript of the gene; or a non-wild-type level of 38646 RNA or protein.

Methods of the invention can be used for prenatal screening or to determine if a subject's offspring will be at risk for a disorder.

In preferred embodiments the method includes determining the structure of a 38646 gene, an abnormal structure being indicative of risk for the disorder.

In preferred embodiments the method includes contacting a sample form the subject with an antibody to the 38646 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

Diagnostic and Prognostic Assays

The presence, level, or absence of 38646 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 38646 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes 38646 protein such that the presence of 38646 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression

of the 38646 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the *38646 genes*; measuring the amount of protein encoded by the 38646 genes; or measuring the activity of the protein encoded by the 38646 genes.

The level of mRNA corresponding to the 38646 gene in a cell can be determined both by in situ and by in vitro formats.

The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length 38646 nucleic acid, such as the nucleic acid of SEQ ID NO: 1, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 38646 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays are described herein.

In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the 38646 genes.

The level of mRNA in a sample that is encoded by 38646 can be evaluated with nucleic acid amplification, e.g., by RT-PCR (U.S. Patent number 4,683,202), ligase chain reaction (Barany, 1991, Proc. Natl. Acad. Sci. USA 88:189-193), self-sustained sequence replication (Guatelli et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh et al., 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi et al., 1988, Bio/Technology 6:1197), rolling circle replication (U.S. Patent number 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5'- or 3'-regions of a 38646 gene (plus and minus strands, respectively, or vice-versa) and contain a short region

in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence between the primers.

For in situ methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the 38646 gene being analyzed.

In another embodiment, the methods include further contacting a control sample with a compound or agent capable of detecting 38646 mRNA, or genomic DNA, and comparing the presence of 38646 mRNA or genomic DNA in the control sample with the presence of 38646 mRNA or genomic DNA in the test sample.

A variety of methods can be used to determine the level of protein encoded by 38646. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled," with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

The detection methods can be used to detect 38646 protein in a biological sample in vitro as well as in vivo. In vitro techniques for detection of 38646 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. In vivo techniques for detection of 38646 protein include introducing into a subject a labeled anti-38646 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting 38646 protein, and comparing the

presence of 38646 protein in the control sample with the presence of 38646 protein in the test sample.

The invention also includes kits for detecting the presence of 38646 in a biological sample. For example, the kit can include a compound or agent capable of detecting 38646 protein or mRNA in a biological sample, and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 38646 protein or nucleic acid.

For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably-labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also includes a buffering agent, a preservative, or a protein-stabilizing agent. The kit can also includes components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples that can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with malexpressed, aberrant or unwanted 38646 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as pain or deregulated cell proliferation.

In one embodiment, a disease or disorder associated with aberrant or unwanted 38646 expression or activity is identified. A test sample is obtained from a subject and 38646 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 38646 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 38646 expression

or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 38646 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent that modulates 38646 expression or activity.

The methods of the invention can also be used to detect genetic alterations in a 38646 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 38646 protein activity or nucleic acid expression, such as a disorder associated with hematopoiesis or an immune disorder. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 38646 protein, or the malexpression of the 38646 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 38646 gene; 2) an addition of one or more nucleotides to a 38646 gene; 3) a substitution of one or more nucleotides of a 38646 gene, 4) a chromosomal rearrangement of a 38646 gene; 5) an alteration in the level of a messenger RNA transcript of a 38646 gene, 6) aberrant modification of a 38646 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a 38646 gene, 8) a non-wild-type level of a 38646 protein, 9) allelic loss of a 38646 gene, and 10) inappropriate post-translational modification of a 38646 protein.

An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE-PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the 38646 gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 38646 gene under conditions such that hybridization and amplification of the 38646 gene occurs (if present), and detecting

the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR can be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh et al., 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi et al., 1988, Bio/Technology 6:1197), or other nucleic acid amplification methods, followed by the detection of the amplified molecules using techniques known to those of skill in the art.

In another embodiment, mutations in a 38646 gene from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined, e.g., by gel electrophoresis, and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (e.g., U.S. Patent number 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in 38646 can be identified by hybridizing a sample to control nucleic acids, e.g., DNA or RNA, by, e.g., two-dimensional arrays, or, e.g., chip based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotides probes (Cronin et al., 1996, Hum. Mutat. 7:244-255; Kozal et al., 1996, Nature Med. 2:753-759). For example, genetic mutations in 38646 can be identified in two-dimensional arrays containing light-generated DNA probes as described (Cronin et al., supra). Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the

characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 38646 gene and detect mutations by comparing the sequence of the sample 38646 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays (1995, Biotechniques 19:448), including sequencing by mass spectrometry.

Other methods for detecting mutations in the 38646 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al., 1985, Science 230:1242; Cotton et al., 1988, Proc. Natl. Acad. Sci. USA 85:4397; Saleeba et al., 1992, Meth. Enzymol. 217:286-295).

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 38646 cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al., 1994, Carcinogenesis 15:1657-1662; U.S. Patent number 5,459,039).

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 38646 genes. For example, single strand conformation polymorphism (SSCP) can be used to detect differences in electrophoretic mobility between mutant and wild-type nucleic acids (Orita et al., 1989, Proc. Natl. Acad. Sci. USA 86:2766; Cotton, 1993, Mutat. Res. 285:125-144; Hayashi, 1992, Genet. Anal. Tech. Appl. 9:73-79). Single-stranded DNA fragments of sample and control 38646 nucleic acids will be denatured and allowed to re-nature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments can be labeled or detected with labeled probes. The sensitivity of the assay can be enhanced by using RNA (rather than DNA), in which the secondary structure

is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al., 1991, Trends Genet 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE; Myers et al., 1985, Nature 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 base pairs of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) Biophys. Chem. 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension (Saiki et al., 1986, Nature 324:163; Saiki et al., 1989, Proc. Natl. Acad. Sci. USA 86:6230).

Alternatively, allele specific amplification technology that depends on selective PCR amplification can be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification can carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; Gibbs et al., 1989, Nucl. Acids Res. 17:2437-2448) or at the extreme 3'-end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner, 1993, Tibtech 11:238). In addition, it can be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al., 1992, Mol. Cell Probes 6:1). It is anticipated that in certain embodiments, amplification can also be performed using Taq ligase for amplification (Barany, 1991, Proc. Natl. Acad. Sci. USA 88:189). In such cases, ligation will occur only if there is a perfect match at the 3'-end of the 5'- sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein can be performed, for example, using pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent

described herein, which can be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a 38646 gene.

Use of 38646 Molecules as Surrogate Markers

The 38646 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 38646 molecules of the invention can be detected, and can be correlated with one or more biological states in vivo. For example, the 38646 molecules of the invention can serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (e.g., with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers can serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (e.g., early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (e.g., an assessment of cardiovascular disease can be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection can be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers have been described (e.g., Koomen et al., 2000, J. Mass. Spectrom. 35:258-264; James, 1994, AIDS Treat. News Arch. 209).

The 38646 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the

presence or activity of the drug in a subject. For example, a pharmacodynamic marker can be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug can be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker can be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug in vivo. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug can be sufficient to activate multiple rounds of marker (e.g., a 38646 marker) transcription or expression, the amplified marker can be in a quantity which is more readily detectable than the drug itself. Also, the marker can be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-38646 antibodies can be employed in an immune-based detection system for a 38646 protein marker, or 38646-specific radiolabeled probes can be used to detect a 38646 mRNA marker. Furthermore, the use of a pharmacodynamic marker can offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers have been described (e.g., U.S. Patent number 6,033,862; Hattis et al., 1991, Env. Health Perspect. 90: 229-238; Schentag, 1999, Am. J. Health-Syst. Pharm. 56 Suppl. 3: S21-S24; Nicolau, 1999, Am, J. Health-Syst. Pharm. 56 Suppl. 3: S16-S20).

The 38646 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (e.g., McLeod et al., 1999, Eur. J. Cancer 35:1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, can be selected. For example, based on the presence or quantity of RNA, or protein (e.g., 38646 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment can be selected that is

optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 38646 DNA can correlate 38646 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

Pharmaceutical Compositions

The nucleic acid and polypeptides, fragments thereof, as well as anti-38646 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. Similarly, compounds which modulate expression of the 38646 gene or activity of 38646 protein can be combined with a pharmaceutically acceptable carrier to form a pharmaceutical composition. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous

preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including an agent in the composition that delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying, which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents and/or adjuvant materials can be included as part

of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder, such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient, such as starch or lactose; a disintegrating agent, such as alginic acid, Primogel™, or corn starch; a lubricant, such as magnesium stearate or Sterotes™; a glidant, such as colloidal silicon dioxide; a sweetening agent, such as sucrose or saccharin; or a flavoring agent, such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser that contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells using monoclonal antibodies directed towards viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to described methods (e.g., U.S. Patent number 4,522,811).

It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds that exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 milligrams per kilogram body weight, preferably about 0.01 to 25 milligrams per kilogram body weight, more preferably about 0.1 to 20 milligrams per kilogram body weight, and even more preferably about 1 to 10 milligrams per kilogram, 2 to 9 milligrams per kilogram, 3 to 8 milligrams per kilogram, 4 to 7 milligrams

per kilogram, or 5 to 6 milligrams per kilogram body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors can influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

For antibodies, the preferred dosage is 0.1 milligrams per kilogram of body weight (generally 10 to 20 milligrams per kilogram). If the antibody is to act in the brain, a dosage of 50 to 100 milligrams per kilogram is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for the lipidation of antibodies is described by Cruikshank et al. (1997, J. AIDS Hum. Retrovir. 14:193).

The present invention encompasses agents that modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including hetero-organic and organo-metallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams

per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

An antibody (or fragment thereof) can be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, and the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety can be a protein or polypeptide possessing a desired

biological activity. Such proteins can include, for example, a toxin such as abrin, ricin A, gelonin, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukins-1, -2, and -6, granulocyte macrophage colony stimulating factor, granulocyte colony stimulating factor, or other growth factors.

Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described in U.S. Patent number 4,676,980.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent number 5,328,470) or by stereotactic injection (e.g., Chen et al., 1994, Proc. Natl. Acad. Sci. USA 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 38646 expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments can be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics," as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype," or "drug response

genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 38646 molecules of the present invention or 38646 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

In one aspect, the invention provides a method for preventing a disease or condition in a subject associated with an aberrant or unwanted 38646 expression or activity, by administering to the subject a 38646 or an agent which modulates 38646 expression, or at least one 38646 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 38646 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 38646 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 38646 aberrance, for example, a 38646, 38646 agonist or 38646 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

Expression of the 38646 gene is positively associated with normal glandular function, and the ability of 38646 protein to induce cytoskeleton organization and intracellular protein or vesicle localization is important for maintenance of normal hormone and protein secretion. Glandular secretion can be inhibited by inhibiting either or both of 38646 gene expression and 38646 protein activity. Agents which can inhibit either 38646 gene expression or 38646 protein activity are preferable delivered in a tissue-specific manner, in order to minimize the effect of the agents on normal 38646 protein function in other tissues. Alternatively, agents which enhance either or both of 38646 gene expression and 38646 protein activity can be administered to glands in which increased secretion is desired.

It is possible that some 38646 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

In addition to the disorders mentioned above, the 38646 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more of the disorders described herein.

As discussed, successful treatment of 38646 disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds, e.g., an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 38646 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, $F(ab')_2$ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof).

Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

Another method by which nucleic acid molecules can be utilized in treating or preventing a disease characterized by 38646 expression is through the use of aptamer molecules specific for 38646 protein. Aptamers are nucleic acid molecules having a tertiary structure that permits them to specifically bind to protein ligands (e.g., Osborne et al., 1997, Curr. Opin.

Chem. Biol. 1:5-9; Patel, 1997, Curr. Opin. Chem. Biol. 1:32-46). Since nucleic acid molecules can in many cases be more conveniently introduced into target cells than therapeutic protein molecules can be, aptamers offer a method by which 38646 protein activity can be specifically decreased without the introduction of drugs or other molecules which can have pluripotent effects.

Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, by administered in instances whereby negative modulatory techniques are appropriate for the treatment of 38646 disorders.

In circumstances wherein injection of an animal or a human subject with a 38646 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against 38646 through the use of anti-idiotypic antibodies (e.g., Herlyn, 1999, Ann. Med. 31:66-78; Bhattacharya-Chatterjee et al., 1998, Cancer Treat. Res. 94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 38646 protein. Vaccines directed to a disease characterized by 38646 expression can also be generated in this fashion.

In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies can be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (e.g., Marasco et al., 1993, Proc. Natl. Acad. Sci. USA 90:7889-7893).

The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or

ameliorate 38646 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

Another example of determination of effective dose for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays can utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate 38646 activity is used as a template, or "imprinting molecule," to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix that contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. Detailed

reviews of this technique appear in the art (Ansell et al., 1996, Curr. Opin. Biotechnol. 7:89-94; Shea, 1994, Trends Polymer Sci. 2:166-173). Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix (e.g., a matrix described in Vlatakis et al., 1993, Nature 361:645-647. Through the use of isotope-labeling, the "free" concentration of compound which modulates the expression or activity of 38646 can be readily monitored and used in calculations of IC_{50}.

Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiber optic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50}. A rudimentary example of such a "biosensor" is discussed in Kriz et al. (1995, Anal. Chem. 67:2142-2144).

Another aspect of the invention pertains to methods of modulating 38646 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a 38646 or agent that modulates one or more of the activities of 38646 protein activity associated with the cell. An agent that modulates 38646 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 38646 protein (e.g., a 38646 substrate or receptor), a 38646 antibody, a 38646 agonist or antagonist, a peptidomimetic of a 38646 agonist or antagonist, or other small molecule.

In one embodiment, the agent stimulates one or 38646 activities. Examples of such stimulatory agents include active 38646 protein and a nucleic acid molecule encoding 38646. In another embodiment, the agent inhibits one or more 38646 activities. Examples of such inhibitory agents include antisense 38646 nucleic acid molecules, anti-38646 antibodies, and 38646 inhibitors. These modulatory methods can be performed in vitro (e.g., by culturing the cell with the agent) or, alternatively, in vivo (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 38646 protein or nucleic acid molecule. In one embodiment, the method involves administering an

agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) 38646 expression or activity. In another embodiment, the method involves administering a 38646 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 38646 expression or activity.

Stimulation of 38646 activity is desirable in situations in which 38646 is abnormally down-regulated and/or in which increased 38646 activity is likely to have a beneficial effect. For example, stimulation of 38646 activity is desirable in situations in which a 38646 is down-regulated and/or in which increased 38646 activity is likely to have a beneficial effect. Likewise, inhibition of 38646 activity is desirable in situations in which 38646 is abnormally up-regulated and/or in which decreased 38646 activity is likely to have a beneficial effect.

Pharmacogenomics

The 38646 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 38646 activity (e.g., 38646 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 38646-associated disorders associated with aberrant or unwanted 38646 activity (e.g., disorders associated with hematopoiesis and immune disorders). In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) can be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician can consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 38646 molecule or 38646 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 38646 molecule or 38646 modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons (e.g., Eichelbaum et al., 1996, Clin. Exp. Pharmacol. Physiol. 23:983-985; Linder et al., 1997, Clin. Chem. 43:254-266). In general, two types of pharmacogenetic conditions can be differentiated.

Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association," relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants). Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high-resolution map can be generated from a combination of some ten million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP can be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that can be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach" can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a 38646 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

Alternatively, a method termed "gene expression profiling," can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed

with a drug (e.g., a 38646 molecule or 38646 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 38646 molecule or 38646 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 38646 genes of the present invention, wherein these products can be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the 38646 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., hematopoietic cells, will become sensitive to treatment with an agent that the unmodified target cells were resistant to.

Monitoring the influence of agents (e.g., drugs) on the expression or activity of a 38646 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 38646 gene expression, protein levels, or up-regulate 38646 activity, can be monitored in clinical trials of subjects exhibiting decreased 38646 gene expression, protein levels, or down-regulated 38646 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 38646 gene expression, protein levels, or down-regulate 38646 activity, can be monitored in clinical trials of subjects exhibiting increased 38646 gene expression, protein levels, or up-regulated 38646 activity. In such clinical trials, the expression or activity of a 38646 gene, and preferably, other genes that have been implicated in, for example, a 38646-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

Other Embodiments

In another aspect, the invention features, a method of analyzing a plurality of capture probes. The method can be used, e.g., to analyze gene expression. The method includes: providing a two-dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence; contacting the array with a 38646, preferably purified, nucleic acid, preferably purified, polypeptide, preferably purified, or antibody, and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the 38646 nucleic acid, polypeptide, or antibody.

The capture probes can be a set of nucleic acids from a selected sample, e.g., a sample of nucleic acids derived from a control or non-stimulated tissue or cell.

The method can include contacting the 38646 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of hybridization can be compared, e.g., to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, e.g., a wild-type, normal, or non-diseased, non-stimulated, sample, e.g., a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, e.g., a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, e.g., a biological fluid, tissue, or cell sample.

The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of 38646. Such methods can be used to diagnose a subject, e.g., to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder. 38646 is associated with hematopoiesis, thus it is useful for evaluating disorders relating to hematopoiesis.

The method can be used to detect SNPs, as described above.

In another aspect, the invention features, a method of analyzing a plurality of probes. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which express 38646 or from a cell or subject in which a 38646 mediated response has been elicited, e.g., by contact of the cell with 38646 nucleic acid or protein, or administration to the cell or subject 38646 nucleic acid or protein; contacting the array with one or more inquiry probe, wherein an inquiry probe can be a nucleic acid, polypeptide, or antibody (which is preferably other than 38646 nucleic acid, polypeptide, or antibody); providing a two-dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which does not express 38646 (or does not express as highly as in the case of the 38646 positive plurality of capture probes) or from a cell or subject which in which a 38646 mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); contacting the array with one or more inquiry probes (which is preferably other than a 38646 nucleic acid, polypeptide, or antibody), and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody.

In another aspect, the invention features, a method of analyzing a plurality of probes or a sample. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality having a unique capture probe, contacting the array with a first sample from a cell or subject which express or malexpress 38646 or from a cell or subject in which a 38646-mediated response has been elicited, e.g., by contact of the cell with 38646 nucleic acid or protein, or administration to the cell or subject 38646 nucleic acid or protein; providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe,

and contacting the array with a second sample from a cell or subject which does not express 38646 (or does not express as highly as in the case of the 38646 positive plurality of capture probes) or from a cell or subject which in which a 38646 mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); and comparing the binding of the first sample with the binding of the second sample. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody. The same array can be used for both samples or different arrays can be used. If different arrays are used the plurality of addresses with capture probes should be present on both arrays.

In another aspect, the invention features a method of analyzing 38646, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 38646 nucleic acid or amino acid sequence, e.g., nucleotide sequence from 38646 or a portion thereof; comparing the 38646 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 38646.

The method can include evaluating the sequence identity between a 38646 sequence and a database sequence. The method can be performed by accessing the database at a second site, e.g., via the internet.

In another aspect, the invention features, a set of oligonucleotides, useful, e.g., for identifying SNPs, or identifying specific alleles of 38646. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, e.g., an SNP or the site of a mutation. In a preferred embodiment, the plurality of oligonucleotides are identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with differential labels, such that an oligonucleotide that hybridizes to one allele provides a signal that is distinguishable from an oligonucleotide that hybridizes to a second allele.

The sequence of a 38646 molecules is provided in a variety of mediums to facilitate use thereof. A sequence can be provided as a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a 38646. Such a manufacture can provide a nucleotide or amino acid sequence, e.g., an open reading frame, in a form which allows

examination of the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form.

A 38646 nucleotide or amino acid sequence can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect™ and Microsoft Word™, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase™, Oracle™, or the like. The skilled artisan can readily adapt any number of data processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. A search is used to identify fragments or regions of the sequences of the invention that match a particular target sequence or target motif.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize

that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Typical sequence lengths of a target sequence are from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, can be of shorter length.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA).

Thus, the invention features a method of making a computer readable record of a sequence of a 38646 sequence that includes recording the sequence on a computer readable matrix. In a preferred embodiment, the record includes one or more of the following: identification of an open reading frame; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5'- end of the translated region; or 5'- and/or 3'-regulatory regions.

In another aspect, the invention features, a method of analyzing a sequence. The method includes: providing a 38646 sequence or record, in computer readable form; comparing a second sequence to the gene name sequence; thereby analyzing a sequence. Comparison can include comparing to sequences for sequence identity or determining if one sequence is included within the other, e.g., determining if the 38646 sequence includes a sequence being compared. In a preferred embodiment, the 38646 or second sequence is stored on a first computer, e.g., at a first site and the comparison is performed, read, or recorded on a second computer, e.g., at a second site. E.g., the 38646 or second sequence can be stored in a public or proprietary database in one computer, and the results of the comparison performed, read, or recorded on a second computer. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full

length amino acid sequence of the protein, or a mature form thereof; the 5'-end of the translated region; or 5'- and/or 3'-regulatory regions.

This invention is further illustrated by the following examples that should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

EXAMPLES

Example 1

Identification and Characterization of Human 38646 cDNA

The human 38646 nucleotide sequence (Figure 1; SEQ ID NO: 1), which is approximately 2561 nucleotides in length including non-translated regions, contains a predicted methionine-initiated coding sequence at about nucleotide residues 115-2413. The coding sequence encodes a 766 amino acid residue protein (SEQ ID NO: 2).

Example 2

Tissue Distribution of 38646 mRNA

Northern blot hybridizations with various RNA samples can be performed under standard conditions and washed under stringent conditions, i.e., 0.2×SSC at 65°C. A DNA probe corresponding to all or a portion of the 38646 cDNA (SEQ ID NO: 1) can be used. The DNA can, for example, be radioactively labeled with 32P-dCTP using the Prime-It™ Kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing mRNA from mouse hematopoietic and endocrine tissues, and cancer cell lines (Clontech, Palo Alto, CA) can be probed in ExpressHyb™ hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

Example 3

Recombinant Expression of 38646 in Bacterial Cells

In this example, 38646 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in E. coli and the fusion polypeptide is isolated and characterized. Specifically, 38646 nucleic acid sequences are fused to GST nucleic acid sequences and this

fusion construct is expressed in E. coli, e.g., strain PEB199. Expression of the GST-38646 fusion construct in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

Example 4

Expression of Recombinant 38646 Protein in COS Cells

To express the 38646 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an E. coli replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire 38646 protein and an HA tag (Wilson et al., 1984, Cell 37:767) or a FLAG tag fused in-frame to its 3'-end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the 38646 DNA sequence is amplified by PCR using two primers. The 5'- primer contains the restriction site of interest followed by approximately twenty nucleotides of the 38646 coding sequence starting from the initiation codon; the 3'-end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the 38646 coding sequence. The PCR amplified fragment and the pcDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the 38646 gene is inserted in the desired orientation. The ligation mixture is transformed into E. coli cells (strains HB101, DH5alpha, SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the 38646-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook et al., (1989, Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). The expression of the 38646 polypeptide is detected by radiolabeling (^{35}S-methionine or ^{35}S-cysteine, available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow et al., 1988, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) using an HA-specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with ^{35}S-methionine (or ^{35}S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 millimolar NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 millimolar Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA-specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the 38646 coding sequence is cloned directly into the polylinker of the pcDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the 38646 polypeptide is detected by radiolabeling and immunoprecipitation using a 38646-specific monoclonal antibody.

Example 5

Expression of the 38646 Gene

Expression of the 38646 gene was assessed in a selection of tissue types using real time quantitative PCR (TAQMAN®) analysis. This data is summarized in Table 1. Relatively high levels of 38646 expression were observed in various brain and nerve tissues, including brain cortex, brain hypothalamus, spinal cord, dorsal root ganglion, and peripheral nerve.

Table 1

Tissue Type	Relative Expression of the 38646 Gene
Normal Artery	35.9
Diseased Aorta	24.0
Normal Vein	5.32
Coronary Smooth Muscle Cell	2.43
Human Umbilical Vein Endothelial Cells	66.5
Hemangioma	35.7
Normal Heart	25.7
Heart - Congestive Heart Failure	42.1
Kidney	83.9
Skeletal Muscle	70.6
Normal Liver	15.3
Normal Small Intestine	10.8
Normal Adipose	18.3
Pancreas	70.8
Primary Osteoblasts	3.17
Normal Bladder - Female	1.02
Normal Adrenal Gland	8.61
Normal Pituitary Gland	30.8
Normal Spinal Cord	63.4
Normal Brain Cortex	187
Normal Brain Hypothalamus	95.1
Nerve	22.3
Dorsal Root Ganglion	35.5
Normal Breast	13.6
Breast Tumor - Infiltrating Ductal Carcinoma	2.26
Normal Ovary	52.6
Ovary Tumor	20.6

Prostate - Benign Prostatic Hyperplasia	4.27
Prostate Adenocarcinoma	11.3
Normal Colon	12.7
Colon Adenocarcinoma	55.0
Normal Lung	8.49
Lung Tumor	13.56
Lung - Chronic Obstructive Pulmonary Disease	4.22
Colon - Inflammatory Bowel Disease	6.94
Synovium	0.94
Normal Tonsil	1.49
Normal Lymph Node	1.64
Liver Fibrosis	15.7
Normal Spleen	1.97
Macrophages	2.78
Progenitor Cells	1.99
(Erythroid, Megakaryocyte, Neutrophil)	
Megakaryocytes	0.66
Activated Peripheral Blood Mononuclear Cells	0.00
Neutrophils	3.81
Erythroid	0.00

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

CLAIMS

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 3;

b) a nucleic acid molecule comprising a fragment of at least 600 nucleotides of the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 3;

c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, wherein the fragment comprises at least 200 contiguous amino acids of SEQ ID NO: 2; and

e) a nucleic acid molecule which encodes a naturally-occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, wherein the nucleic acid molecule hybridizes with a nucleic acid molecule comprising SEQ ID NO: 1, SEQ ID NO: 3 or a complement thereof, under stringent conditions.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 3; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell that contains the nucleic acid molecule of claim 1.

6. The host cell of claim 5, wherein the host cell is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 1, SEQ ID NO: 3, or a complement thereof.

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes with a nucleic acid molecule comprising SEQ ID NO: 1, SEQ ID NO: 3, or a complement thereof under stringent conditions; and

c) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, wherein the fragment comprises at least 200 contiguous amino acids of SEQ ID NO: 2.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO: 2.

10. The polypeptide of claim 8, further comprising heterologous amino acid sequences.

11. An antibody that selectively binds with a polypeptide of claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO: 2;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO: 2, wherein the fragment comprises at least 200 contiguous amino acids of SEQ ID NO: 2; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes with a nucleic acid molecule comprising SEQ ID NO: 1, SEQ ID NO: 3 or a complement thereof, under stringent conditions;

the method comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds with a polypeptide of claim 8; and

b) determining whether the compound binds with the polypeptide in the sample.

14. The method of claim 13, wherein the compound that binds with the polypeptide is an antibody.

15. A kit comprising a compound that selectively binds with a polypeptide of claim 8 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes with the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds with a nucleic acid molecule in the sample.

17. The method of claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound that selectively hybridizes with a nucleic acid molecule of claim 1 and instructions for use.

19. A method for identifying a compound which binds with a polypeptide of claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and

b) determining whether the polypeptide binds with the test compound.

20. The method of claim 19, wherein the binding of the test compound with the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay; and

c) detection of binding using an assay for 38646-mediated signal transduction.

21. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds with the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

22. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

a) contacting a polypeptide of claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```
TGCAAGTCCAAAGCCACAAGGTGCCTCCAGAAGCCATTACCCCTGCAGAATTCACCTTCGTCCAATATACACCAAACCC

                                        M   E   E   I   K   P   A   S   A   S   C    11
CCAGGCATAAAGCTTTACCTAGTGCAAAACCAAGG ATG GAG GAA ATT AAA CCT GCC TCT GCT TCT TGT    33

 V   S   K   E   K   P   S   K   V   S   D   L   I   S   R   F   E   G   G   S    31
GTC TCA AAA GAA AAA CCC AGT AAG GTA TCA GAT CTC ATC AGT CGC TTT GAA GGA GGC AGC    93

 S   L   S   N   Y   S   D   L   K   K   E   S   A   V   N   L   N   A   P   R    51
TCA TTA TCA AAT TAT AGT GAT TTG AAG AAA GAG TCT GCT GTG AAC CTA AAT GCT CCT AGA   153

 T   P   G   R   H   G   L   T   T   T   P   Q   Q   K   L   L   S   Q   H   L    71
ACC CCA GGA AGG CAT GGA TTG ACA ACC ACA CCT CAA CAA AAA CTC CTC TCC CAG CAC TTG   213

 P   Q   R   Q   G   N   D   T   D   K   T   Q   G   A   Q   T   C   V   A   N    91
CCA CAG AGG CAG GGA AAT GAT ACA GAT AAG ACT CAG GGT GCA CAG ACT TGT GTG GCC AAC   273

 G   V   M   A   A   Q   N   Q   M   E   C   E   E   E   K   A   A   T   L   S   111
GGT GTA ATG GCA GCA CAA AAC CAG ATG GAA TGT GAG GAG GAG AAA GCT GCC ACT CTT AGC   333

 S   D   T   S   I   Q   A   S   E   P   L   L   D   T   H   I   V   N   G   E   131
TCA GAT ACT TCT ATT CAA GCT TCT GAA CCC TTG CTT GAT ACG CAC ATA GTG AAT GGA GAA   393

 R   D   E   T   A   T   A   P   A   S   P   T   T   D   S   C   D   G   N   A   151
AGA GAT GAA ACT GCC ACA GCT CCT GCA TCA CCC ACA ACA GAT AGC TGT GAT GGA AAT GCT   453
```

Fig. 1A

```
 S   D   S   S   Y   R   T   P   G   I   G   P   V   L   P   L   E   E   R   G   171
TCT GAC AGT AGC TAC AGG ACT CCA GGC ATA GGC CCA GTG CTC CCC CTA GAA GAA AGA GGG   513

 A   E   T   E   T   K   V   Q   E   R   E   N   G   E   S   P   L   E   L   E   191
GCA GAA ACA GAA ACC AAG GTA CAA GAG AGG GAA AAT GGG GAA AGC CCT CTG GAA CTG GAG   573

 Q   L   D   Q   H   H   E   M   K   E   T   N   E   Q   K   L   H   K   I   A   211
CAG CTG GAC CAG CAC CAT GAG ATG AAG GAG ACT AAT GAG CAA AAA CTT CAC AAA ATA GCC   633

 N   E   L   L   L   T   E   R   A   Y   V   N   R   L   D   L   L   D   Q   V   231
AAT GAA CTT TTG CTT ACT GAA AGA GCT TAT GTC AAC CGA CTT GAC CTC TTA GAT CAG GTA   693

 F   Y   C   K   L   L   E   E   A   N   R   G   S   F   P   A   E   M   V   N   251
TTT TAT TGC AAA CTG TTG GAA GAA GCA AAC CGA GGC TCG TTT CCA GCA GAG ATG GTG AAT   753

 K   I   F   S   N   I   S   S   I   N   A   F   H   S   K   F   L   L   P   E   271
AAA ATC TTT TCT AAT ATT TCA TCA ATA AAT GCC TTC CAT AGT AAA TTC CTC TTG CCA GAG   813

 L   E   K   R   M   Q   E   W   E   T   T   P   R   I   G   D   I   L   Q   K   291
CTG GAG AAA CGA ATG CAA GAA TGG GAA ACT ACT CCT AGA ATT GGA GAC ATC CTT CAG AAA   873

 L   A   P   F   L   K   M   Y   G   E   Y   V   K   G   F   D   N   A   M   E   311
TTG GCA CCA TTC CTT AAG ATG TAT GGA GAA TAT GTG AAA GGA TTT GAT AAT GCA ATG GAA   933

 L   V   K   N   M   T   E   R   I   P   Q   F   K   S   V   V   E   E   I   Q   331
TTG GTT AAA AAC ATG ACA GAA CGT ATT CCC CAG TTC AAA TCA GTG GTT GAA GAA ATT CAG   993
```

Fig. 1B

```
 K   Q   K   I   C   G   S   L   T   L   Q   H   H   M   L   E   P   V   Q   R   351
AAA CAG AAA ATC TGT GGG AGC TTA ACT TTG CAG CAT CAC ATG CTA GAA CCT GTT CAG CGG 1053

 I   P   R   Y   E   M   L   L   K   D   Y   L   R   K   L   P   P   D   S   L   371
ATT CCC CGG TAT GAG ATG CTC CTT AAG GAC TAT CTA AGG AAA TTG CCT CCT GAT TCC CTG 1113

 D   W   N   D   A   K   K   S   L   E   I   I   S   T   A   A   S   H   S   N   391
GAC TGG AAT GAT GCT AAA AAA TCA CTT GAA ATT ATA TCT ACA GCA GCA AGC CAT TCT AAT 1173

 S   A   I   R   K   M   E   N   L   K   K   L   L   E   I   Y   E   M   L   G   411
AGT GCA ATA AGG AAA ATG GAG AAC CTA AAG AAA CTC TTA GAG ATT TAT GAA ATG TTG GGA 1233

 E   E   E   D   I   V   N   P   S   N   E   L   I   K   E   G   Q   I   L   K   431
GAA GAA GAA GAC ATT GTA AAC CCT TCA AAT GAA CTA ATA AAA GAA GGA CAG ATC CTC AAA 1293

 L   A   A   R   N   T   S   A   Q   E   R   Y   L   F   L   F   N   N   M   L   451
CTA GCT GCT CGG AAC ACT TCA GCA CAA GAA CGC TAC CTT TTC TTA TTC AAC AAC ATG TTG 1353

 L   Y   C   V   P   K   F   S   L   V   G   S   K   F   T   V   R   T   R   V   471
CTG TAC TGT GTG CCC AAA TTC AGC TTG GTA GGC TCT AAA TTC ACA GTT CGA ACC AGG GTT 1413

 G   I   D   G   M   K   I   V   E   T   Q   N   E   E   Y   P   H   T   F   Q   491
GGC ATT GAT GGA ATG AAA ATT GTA GAG ACT CAA AAT GAA GAA TAT CCA CAT ACT TTC CAG 1473

 V   S   G   K   E   R   T   L   E   L   Q   A   S   S   A   Q   D   K   E   E   511
GTG TCT GGG AAA GAG AGA ACA CTG GAA CTG CAG GCC AGT TCT GCG CAA GAC AAA GAA GAA 1533
```

Fig. 1C

```
 W   I   K   A   L   Q   E   T   I   D   A   F   H   Q   R   H   E   T   F   R   531
TGG ATC AAG GCC CTT CAA GAA ACC ATC GAT GCT TTT CAT CAA AGG CAT GAA ACC TTC AGA 1593

 N   A   I   A   K   D   N   D   I   H   S   E   V   S   T   A   E   L   G   K   551
AAT GCA ATT GCA AAG GAT AAT GAC ATT CAC TCA GAG GTT TCT ACT GCT GAG CTA GGG AAA 1653

 R   A   P   R   W   I   R   D   N   E   V   T   M   C   M   K   C   K   E   P   571
AGA GCC CCA AGA TGG ATC CGA GAT AAT GAA GTG ACA ATG TGT ATG AAA TGT AAA GAA CCT 1713

 F   N   A   L   T   R   R   R   H   H   C   R   A   C   G   Y   V   V   C   W   591
TTC AAT GCA CTG ACA CGA AGG AGG CAT CAT TGT CGA GCA TGT GGA TAT GTG GTT TGT TGG 1773

 K   C   S   D   Y   K   A   Q   L   E   Y   D   G   G   K   L   S   K   V   C   611
AAA TGC TCC GAC TAC AAA GCT CAA CTT GAA TAT GAT GGT GGT AAA TTG AGC AAA GTT TGT 1833

 K   D   C   Y   Q   I   I   S   G   F   T   D   S   E   E   K   K   R   K   G   631
AAA GAC TGT TAT CAA ATC ATA AGT GGA TTC ACA GAC AGT GAA GAA AAA AAA AGA AAA GGA 1893

 I   L   E   I   E   S   A   E   V   S   G   N   S   V   V   C   S   F   L   Q   651
ATT TTA GAG ATT GAA TCA GCA GAA GTA TCT GGA AAC AGT GTG GTG TGC AGC TTT CTT CAG 1953

 Y   M   E   K   S   K   P   W   Q   K   A   W   C   V   I   P   K   Q   D   P   671
TAT ATG GAG AAG TCA AAA CCT TGG CAG AAA GCT TGG TGT GTG ATC CCC AAG CAA GAC CCT 2013

 L   V   L   Y   M   Y   G   A   P   Q   D   V   R   A   Q   A   T   I   P   L   691
CTT GTG CTG TAC ATG TAT GGT GCC CCC CAG GAC GTC AGA GCC CAG GCC ACC ATT CCA CTT 2073
```

Fig. 1D

```
 L   G   Y   V   V   D   E   M   P   R   S   A   D   L   P   H   S   F   K   L   711
CTG GGC TAT GTG GTG GAT GAA ATG CCA AGG AGC GCA GAC CTG CCA CAC AGT TTC AAA CTG 2133

 T   Q   S   K   S   V   H   S   F   A   A   D   S   E   E   L   K   Q   K   W   731
ACC CAG TCT AAG TCC GTG CAC AGC TTT GCT GCA GAC AGT GAG GAA CTG AAG CAG AAG TGG 2193

 L   K   V   I   L   L   A   V   T   G   E   T   P   G   G   P   N   E   H   P   751
CTG AAA GTC ATC CTT TTA GCT GTC ACA GGT GAG ACA CCA GGT GGT CCA AAT GAG CAT CCA 2253

 A   T   L   D   D   H   P   E   P   K   K   K   S   E   C   *                   766
GCC ACC TTG GAT GAT CAT CCT GAA CCT AAG AAA AAA TCA GAA TGC TGA                   2301

ACTCCTCCAGGACCAGCCATGGTGTGGAGGTCTCAGGACTTACAGCTCAAGACATTCCCAGCTCTTCTTACACATCTGC
TAGCACTTTATGTTGAAAAATATAGGCCCATAAATGCAAAAAAAAAAAAAAAAAGGGCGGCCGAAA
```

Fig. 1E

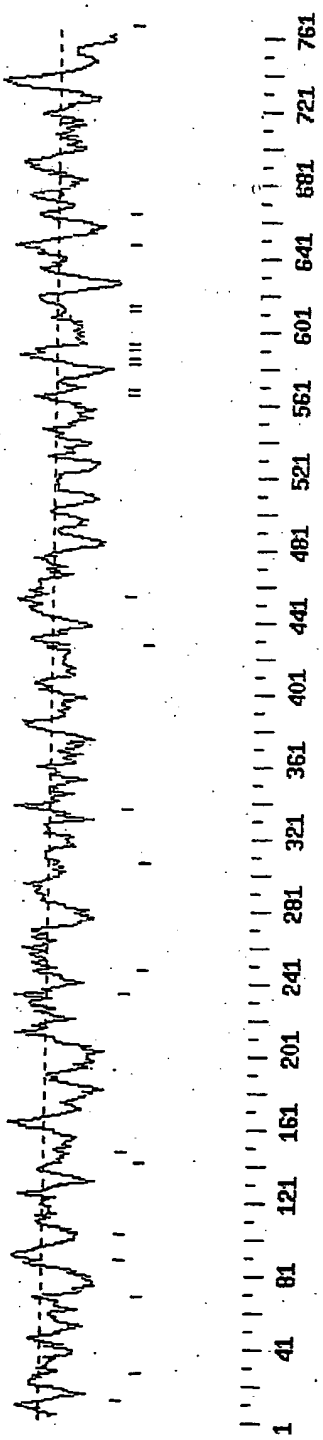
1
41
81
121
161
201
241
281
321
361
401
441
481
521
561
601
641
681
721
761

Fig. 2

```
Frabin MEESNPAPTSCASKGKHSKVSDLISHFEGGSVLSSYTDVQKDSTMNLNIPQTPRQHGLTSTTPQ
       :::  ::  :: :: : :::::::: ::::: :: : :  : :  ::: : ::  :::: :  :
 38646 MEEIKPASASCVSKEKPSKVSDLISRFEGGSSLSNYSDLKKESAVNLNAPRTPGRHGLTTTPQQ

Frabin KLPSHKSPQKQEKDSDQNQGQHGCLANGVAAAQSQMECETEKEAALSPETDTQTAAASPDAHVL
       :: :   :: :  : :  ::     :::: ::: ::::: :: : ::  :  :       : :
 38646 KLLSQHLPQRQGNDTDKTQGAQTCVANGVMAAQNQMECEEEKAATLSSDTSIQASEPLLDTHIV

Frabin NGVRNETTTDSASSVTNSHDENACDSSCRTQGTDLGLPSKEGEPVIEAELQERENGLSTEGLNP
       :: : :: :  ::  : : : :: ::: :: :   ::  :    :  ::::: :   :
 38646 NGERDETATAPASPTTDSCDGNASDSSYRTPGIGPVLPLEERGAETETKVQERENGESPLELEQ

Frabin LDQHHEVKETNEQKLHKIATELLLTERAYVSRLNLLDQVFYCKLLEEANRGSFPAEMVNKIFSN
       :::::: :::::::::::: :::::::::: :: ::::::::::::::::::::::::::::::
 38646 LDQHHEMKETNEQKLHKIANELLLTERAYVNRLDLLDQVFYCKLLEEANRGSFPAEMVNKIFSN

Frabin ISSINAFHSKFLLPELEKRMQEWETTPRIGDILQKLAPFLKMYGEYVKGFDNAVELVKNMTERV
       :::::::::::::::::::::::::::::::::::::::::::::::::::: :::::::::
 38646 ISSINAFHSKFLLPELEKRMQEWETTPRIGDILQKLAPFLKMYGEYVKGFDNAMELVKNMTERI

Frabin PQFKSVTEEIQKQKICGSLTLQHHMLEPIQRIPRYEMLLKDYLKKLSPDAPDWNDAKKSLEIIS
       :::::: :::::::::::::::::::: :::::::::::::: :: ::  ::::::::::::
 38646 PQFKSVVEEIQKQKICGSLTLQHHMLEPVQRIPRYEMLLKDYLRKLPPDSLDWNDAKKSLEIIS
```

Fig. 3A

```
Frabin TAASHSNSAIRKMENLKKLLEIYEMLGEEEDIVNPSNELIKEGQILKLAARNTSAQERYLFLFN
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
 38646 TAASHSNSAIRKMENLKKLLEIYEMLGEEEDIVNPSNELIKEGQILKLAARNTSAQERYLFLFN

Frabin NMLLYCVPRFSLVGSKFTVRTRVGIDGMKIVETHNEEYPHTFQVSGKERTLELQASSEQDKEEW
       :::::::: ::::::::::::::::::::::::: :::::::::::::::::::::::::: ::::::
 38646 NMLLYCVPKFSLVGSKFTVRTRVGIDGMKIVETQNEEYPHTFQVSGKERTLELQASSAQDKEEW

Frabin IKALQESIDAFHQRHETFRNAIAKENDIPLEVSTAELGKRAPRWIRDNEVTMCMKCKESFNALT
       :::::: ::::::::::::::::: :::  :::::::::::::::::::::::::::::: :::::
 38646 IKALQETIDAFHQRHETFRNAIAKDNDIHSEVSTAELGKRAPRWIRDNEVTMCMKCKEPFNALT

Frabin RRRHHCRACGHVVCWKCSDYKAQLEYDGGRLNKVCKDCYQIMSGFAESEEKKRRGILEIESAEV
       :::::::::: :::::::::::::::::: : ::::::::: ::: ::::::: ::::::::::
 38646 RRRHHCRACGYVVCWKCSDYKAQLEYDGGKLSKVCKDCYQIISGFTDSEEKKRKGILEIESAEV

Frabin SGNSEVCSFLQYMEKSKPWQKIWCVIPKQDPLVLYMYGAPQDVRAQATIPLLGYIVDDMPKSAD
       :::: ::::::::::::::: ::::::::::::::::::::::::::::::::: :: :: :::
 38646 SGNSVVCSFLQYMEKSKPWQKAWCVIPKQDPLVLYMYGAPQDVRAQATIPLLGYVVDEMPRSAD

Frabin LPHSFKLTQSKSVHSFAADSEELKQKWLKIILLAVTGETPDGPSEHLDTLDNLPGFKEKSEC
       :::::::::::::::::::::::::::::: :::::::::: :: ::  ::: : :: ::::
 38646 LPHSFKLTQSKSVHSFAADSEELKQKWLKVILLAVTGETPGGPNEHPATLDDHPEPKKKSEC
```

Fig. 3B

```
mFGD1       MHGHRVPGGPGPSDPERSAANTPGAAPLACADSDPGALEPGLPVSRGSGTALGGPLDPQF
hFGD1       MHGHRAPGGRRAFGARTPGHEPAGAAPPACADSDPGASEPGLLARRGSGSALGGPLDPQF
mFGD3       ------------------------------------------------------------
rFrabin     ------------------------------------------------------------
mFGD2       ------------------------------------------------------------

38646       ------------------------------------------------------------


mFGD1       VGPSDASLGAPPSSRVLPCGPSPQHHRALRFSYHLEGSQPRPGLHQGNRILVKSLSLDPG
hFGD1       VGPSDTSLGAAPGHRVLPCGPSPQHHRALRFSYHLEGSQPRPGLHQGNRILVKSLSLDPG
mFGD3       ------------------------------------------------------------
rFrabin     ------------------------------------------------------------
mFGD2       ------------------------------------------------------------

38646       ------------------------------------------------------------


mFGD1       QSLEPHPEGPQRLRSDPGPPTEIPGPRPSPLKRAPGPKPQVPPKPSYLQMPRVLPPPEPI
hFGD1       QSLEPHPEGPQRLRSDPGPPTETPSQRPSPLKRAPGPKPQVPPKPSYLQMPRMPPPLEPI
mFGD3       ------------------------------------------------------------
rFrabin     ------------------------------------------------------MEESN
mFGD2       ------------------------------------------------------------

38646       ------------------------------------------------------MEEIK
```

Fig. 4A

```
mFGD1      PPPPSRPLPADPRVAKGLVPRAEASTSSAAVSSLIEKFEREPVIVASDRPAPGPCPV-PP
hFGD1      PPPPSRPLPADPRVGKGLAPRAEASPSSAAVSSLIEKFEREPVIVASDRPVPGPSPG-PP
mFGD3      ----------------MELGRSSSTPQEEAIS---------PLGVLGTGPSSSPLGK-LQ
rFrabin    PAPTSCASKGKHSKVSDLISHFEGGSVLSSYTDVQ-KDSTMNLNIPQTPRQHGLTSTTPQ
mFGD2      ------------------------------------------------------------

38646      PASASCVSKEKPSKVSDLISRFEGGSSLSNYSDLK-KESAVNLNAPRTPGRHGLTTTPQQ


mFGD1      EPAMLPQPPPQPTGSQLPEGEASRCLFLLAPGPRDGEKVPNRDSGIDSISSPSNSE-ETC
hFGD1      EPVMLPQPTSQPPVPQLPEGEASRCLFLLAPGPRDGEKVPNRDSGIDSISSPSNSE-ETC
mFGD3      ALPIGPGAHRGAHSSSAPAGDSS------TREPSGAMKIPNRDSGIDSPSSSVASENFPC
rFrabin    KLPSHKSPQKQEKDSDQNQGQHGCLANGVAAAQSQMECETEKEAALSPETDTQTAAASPD
mFGD2      ----------------------------MERACEKQDSVCNLVAVFENNRTPGEAPGSHS
                                                   :  : :.    .  :
38646      KLLSQHLPQRQGNDTDKTQGAQTCVANGVMAAQNQMECEEEKAATLSSDTSIQASEPLLD


mFGD1      FVSDDGPPIHSLCPGPPALASMPVALADPHRPGSQEVDS--DLEEEE-EEEEEEKEREI-
hFGD1      FVSDDGPPSHSLCPGPPALASVPVALADPHRPGSQEVDS--DLEEEDDEEEEEEKDREI-
mFGD3      EESSEGSP-------SPAILGLPSETASDSR-VPQ--DN--PQEEEDSGVGEEPDPK---
rFrabin    AHVLNGVRNETTTDSASSVTNSHDENACDSSCRTQGTDLGLPSKEGEPVIEAELQERENG
mFGD2      LEDQPHIPEHQLSLSPEPWEAPPVKEALKSEFRPVSRTY---LSSLKNKLSSGAWRRSC-
                          . .        *      .         .. .         :
38646      THIVNGERDETATAPASPTTDSCDGNASDSSYRTPGIGPVLPLEERGAETETKVQERENG
```

Fig. 4B

```
mFGD1          ----PVPPMERQESVELTVQ---QKVFHIANELLQTEKAYVSRLHLLDQVFCARLLEEAR
hFGD1          ----PVPLMERQESVELTVQ---QKVFHIANELLQTEKAYVSRLHLLDQVFCARLLEEAR
mFGD3          -----VTLFRPQEDVSLTQCSDPQKLLHIAQELLHTEEAYVKRLHLLDQVFCTKLTEAG-
rFrabin        LSTEGLNPLDQHHEVKETNE---QKLHKIATELLLTERAYVSRLNLLDQVFYCKLLEEAN
mFGD2          -----QPGVSPGPETQEPEE------KRVVRELLETEQAYVARLHLLDQVFFQELLREAG
                     .    ... .         ::. *** **.*** **:******  .* . .
38646          ESPLELEQLDQHHEMKETNE---QKLHKIANELLLTERAYVNRLDLLDQVFYCKLLEEAN

mFGD1          NRSSFPADVVHGIFSNICSIYCFHQQFLLPELEKRM-EEWDRYPRIGDILQKLAPFLKMY
hFGD1          NRSSFPADVVHGIFSNICSIYCFHQQFLLPELEKRM-EEWDRYPRIGDILQKLAPFLKMY
mFGD3          ----IPLEVTTGIFSNISSIYRFHGQFLLPELQKRITEEWDTNPRLGDILQKLAPFLKMY
rFrabin        -RGSFPAEMVNKIFSNISSINAFHSKFLLPELEKRM-QEWETTPRIGDILQKLAPFLKMY
mFGD2          RSKAFPEDVVKLIFSNISSIYRFHAQFFLPELQRRV-DDWAATPRIGDVIQKLAPFLKMY
                   :* ::.  *****.**  ** :*:****::*: ::*   **:**::**********
38646          -RGSFPAEMVNKIFSNISSINAFHSKFLLPELEKRM-QEWETTPRIGDILQKLAPFLKMY

mFGD1          GEYVKNFDRAVELVNTWTERSTQFKVIIHEVQKEEACRNLTLQHHMLEPVQRIPRYELLL
hFGD1          GEYVKNFDRAVELVNTWTERSTQFKVIIHEVQKEEACGNLTLQHHMLEPVQRIPRYELLL
mFGD3          GEYVKNFDRAMGLVSTWTQRSPQFKDVIHTIQKQEVCGNLTLQHHMLEPVQRVPRYELLL
rFrabin        GEYVKGFDNAVELVKNMTERVPQFKSVTEEIQKQKICGSLTLQHHMLEPIQRIPRYEMLL
mFGD2          SEYVKNFERAAELLATWMDKSQPFQEVVTRIQCSEASSSLTLQHHMLEPVQRIPRYELLL
               .****.*:.*  *: .  ::   *: :   :* .: . .**********:**:****:**
38646          GEYVKGFDNAMELVKNMTERIPQFKSVVEEIQKQKICGSLTLQHHMLEPVQRIPRYEMLL
```

Fig. 4C

```
mFGD1          KDYLLKLPHGSPDSKDAQKSLELIATAAEHSNAAIRKMERMHKLLKVYELLGGEEDIVSP
hFGD1          KDYLLKLPHGSPDSKDAQKSLELIATAAEHSNAAIRKMERMHKLLKVYELLGGEEDIVSP
mFGD3          KDYLKRLPRDAPDRKDAERSLELISTAADHSNAAIRKMEKMHKLLEVYEQLGGEEDIVNP
rFrabin        KDYLKKLSPDAPDWNDAKKSLEIISTAASHSNSAIRKMENLKKLLEIYEMLGEEEDIVNP
mFGD2          KEYVQKLPAQAPDLEDAQRALDMIFSAAQHSNAAIAEMERLQGLWDVYQRLGLEDDIVDP
               *:*: :*.  :** :**:::*::* :**.***:** :**.:: * .:*: ** *:***.*
38646          KDYLRKLPPDSLDWNDAKKSLEIISTAASHSNSAIRKMENLKKLLEIYEMLGEEEDIVNP

mFGD1          TKELIKEGHILKLSAKNGTTQDRYLILFNDRLLYCVPRLRLLGQKFTVRARIDVDGMELK
hFGD1          TKELIKEGHILKLSAKNGTTQDRYLILFNDRLLYCVPRLRLLGQKFSVRARIDVDGMELK
mFGD3          ANELIKEGSIQKLSAKNGTTQDRHLFLFNNVMLYCVPKLRLMGQKLSVREKMDISDLQVQ
rFrabin        SNELIKEGQILKLAARNTSAQERYLFLFNNMLLYCVPRFSLVGSKFTVRTRVGIDGMKIV
mFGD2          SNTLLREGPVLKISFRRSDPMERYLVLFNNMLLYCVPRVLQVGAQFQVRTRIDVAGMKVR
               :: *::** : *:: :.  . :*:*.***: :*****:.  :* :: ** :..: .:::
38646          SNELIKEGQILKLAARNTSAQERYLFLFNNMLLYCVPKFSLVGSKFTVRTRVGIDGMKIV

mFGD1          ESSNLNMPRTFLVSGKQRSLELQARTEEEKKDWVQAINSTLLKHEQTLETFKLLNSTNRD
hFGD1          ESSNLNLPRTFLVSGKQRSLELQARTEEEKKDWVQAINSTLLKHEQTLETFKLLNSTNRE
mFGD3          DIVKPNAACTFIITGRKRSLELQTRTEEEKKEWIQVIQATVEKHKQKSETFRAFGGACSQ
rFrabin        ETHNEEYPHTFQVSGKERTLELQASSEQDKEEWIKALQESIDAFHQRHETFR---NAIAK
mFGD2          ELTDAEFPHSFLVSGKQRTLELQARSRDEMVSWMQACQAAIDQVEKRSETFKAAVQGPQG
               :  . : . :* ::*::*:****: :.::  .*::. : ::   .:  ***:
38646          ETQNEEYPHTFQVSGKERTLELQASSAQDKEEWIKALQETIDAFHQRHETFR---NAIAK
```

Fig. 4D

```
mFGD1          DED---TP-----------PN--------SPNVDLGKRAPTPIREKEVTMCMRCQEPFNS
hFGD1          DED---TP-----------PN--------SPNVDLGKRAPTPIREKEVTMCMRCQEPFNS
mFGD3          DEEPTLSPDQPVMSTSSVEPAGVADSNGGTPGIESRKSSSKTRRDKEKPGCKSCGETFNS
rFrabin        END---IP-------------------LEVSTAELGKRAPRWIRDNEVTMCMKCKESFNA
mFGD2          DTQ---EP--------------------KPQVEELGLRAPQWVRDKMVTMCMRCQEPFNA
               : :    *                          :    :.   *::  . *  * *.**:
38646          DND---IH-------------------SEVSTAELGKRAPRWIRDNEVTMCMKCKEPFNA

mFGD1          ITKRRHHCKACGHVVCGKCSEFRARLIYDNNRSNRVCTDCYVALHGAPGSSPACSQHTPQ
hFGD1          ITKRRHHCKACGHVVCGKCSEFRARLVYDNNRSNRVCTDCYVALHGVPGSSPACSQHTPQ
mFGD3          ITKRRYRCKLCGEVICRKCSEFKAE----NSKQSRVCRECFLEEPLVP---PSPSSETPT
rFrabin        LTRRRHHCRACGHVVCWKCSDYKAQLEYDGGRLNKVCKDCYQIMSGFAE-----SEEK--
mFGD2          LTRRRHHCRACGYVVCAKCSDYRAELKYDSNRPNRVCLTCYTFLTGNVL--P---QGKED
               :*:**::*: ** *:* ***:::*.    ..: .:**  *:                . .
38646          LTRRRHHCRACGYVVCWKCSDYKAQLEYDGGKLSKVCKDCYQIISGFTD-----SEEK--

mFGD1          RRRSILEKQASVAAENSVICSFLHYMEKGG-KGWHKAWFVVPENEPLVLYIYGAPQDVKA
hFGD1          RRRSILEKQASVAAENSVICSFLHYMEKGG-KGWHKAWFVVPENEPLVLYIYGAPQDVKA
mFGD3          ELKQNAEKPPSVDPRPSLLCGTLNLSDDG--TTWNEVWAAIPESDPQVLDLLAGSQAGRL
rFrabin        KRRGILEIESAEVSGNSEVCSFLQYMEKS--KPWQKIWCVIPKQDPLVLYMYGAPQDVRA
mFGD2          KRRGILEKEASAAPEQSLVCSFLQLIGDKCSRSLPRSWCVIPRDDPLVLYVYAAPQDTKA
               . :   *  .:  .  * :*. *:   .        . * .:*..:* ** : ...*  :
38646          KRKGILEIESAEVSGNSVVCSFLQYMEKS--KPWQKAWCVIPKQDPLVLYMYGAPQDVRA
```

Fig. 4E

```
mFGD1          QRSLPLIGFEVGP-PEAGERPDRRHVFKITQ------------SHLSWYFSPETEELQRR
hFGD1          QRSLPLIGFEVGP-PEAGERPDRRHVFKITQ------------SHLSWYFSPETEELQRR
mFGD3          LYSIPLSGCNITM-PDPEEGLEAGCAWKLHQ------------GSQTWWLSAPSTKLQQC
rFrabin        QATIPLLGYIVD---DMPKSADLPHSFKLTQ------------SKSVHSFAADSEELKQK
mFGD2          HTSIPLLGYQVISGPQGTSGFPAATVRRAVHLQGRVCGAAGPLGDSYQACASGTPEDLTK
                 ::** *  :    :  .        :  :             .     :. : :
38646          QATIPLLGYVVD---EMPRSADLPHSFKLTQ------------SKSVHSFAADSEELKQK

mFGD1          WMAVLGRAGRGDTFC----PGPT----------LSEDKE---MEETPVAASGATAEPPEA
hFGD1          WMAVLGRAGRGDTFC----PGPT----------LSEDRE---MEEAPVAALGATAEPPES
mFGD3          WLKALGTAVHGDTAG----DRPG----------ASQ----------PQAPAGTDT-P---
rFrabin        WLKIILLAVTGETP-----DGP-----------SE----------HLDTLDNLPGPKEK
mFGD2          KMCLTEPAASCSSRVHDSLPRPTPWFYHFTPSWATPDPPDVYTETIHPDSVSSRHRPFPS
                :    *   .:       *             :            . .    *
38646          WLKVILLAVTGETP-----GGP-----------NE----------HPATLPDHPEPKKK

mFGD1          SQTRDKT-----------
hFGD1          PQTRDKT-----------
mFGD3          ------------------
rFrabin        SEC---------------
mFGD2          GRYHQVSQLAGEGADIPGS

38646          SEC---------------
```

Fig. 4F

```
848         SSLSNYSDLKKESAVNLNAPRTPGRHGLTTTPQQKLLSQHLPQRQGNDTDKTQGAQTCVA 384
38646   12  SSLSNYSDLKKESAVNLNAPRTPGRHGLTTTPQQKLLSQHLPQRQGNDTDKTQGAQTCVA 71

848         NGVMAAQNQMECEEEKAATLSSDTSIQASEPLLDTHIVNGERDETATAPASPTTDSCDGN 564
38646   72  NGVMAAQNQMECEEEKAATLSSDTSIQASEPLLDTHIVNGERDETATAPASPTTDSCDGN 131

848         ASDSSYRTPGIGPVLPLEERGAETETKVQERENGESPLELEQLDQHHEMKETNE 726
38646  132  ASDSSYRTPGIGPVLPLEERGAETETKVQERENGESPLELEQLDQHHEMKVEHE 185
```

Fig. 5

SEQUENCE LISTING

<110> Millennium Pharmaceuticals, Inc.

<120> 38646, A Novel Guanine Nucleotide Exchange Factor

<130> 10147-44WO

<140> Not Yet Assigned
<141> 2001-09-10

<150> 60/231,089
<151> 2000-09-08

<160> 9

<170> PatentIn version 3.0

<210> 1
<211> 2560
<212> DNA
<213> Homo sapiens

<400> 1
tgcaagtcca aagccacaag gtgcctccag aagccattac ccctgcagaa ttcaccttcg 60

tccaatatac accaaacccc caggcataaa gctttaccta gtgcaaaacc aaggatggag 120

gaaattaaac ctgcctctgc ttcttgtgtc tcaaaagaaa aacccagtaa ggtatcagat 180

ctcatcagtc gctttgaagg aggcagctca ttatcaaatt atagtgattt gaagaaagag 240

tctgctgtga acctaaatgc tcctagaacc ccaggaaggc atggattgac aaccacacct 300

caacaaaaac tcctctccca gcacttgcca cagaggcagg gaaatgatac agataagact 360

cagggtgcac agacttgtgt ggccaacggt gtaatggcag cacaaaacca gatggaatgt 420

gaggaggaga aagctgccac tcttagctca gatacttcta ttcaagcttc tgaacccttg 480

cttgatacgc acatagtgaa tggagaaaga gatgaaactg ccacagctcc tgcatcaccc 540

acaacagata gctgtgatgg aaatgcttct gacagtagct acaggactcc aggcataggc 600

ccagtgctcc ccctagaaga aagaggggca gaaacagaaa ccaaggtaca agagagggaa 660

aatggggaaa gccctctgga actggagcag ctggaccagc accatgagat gaaggagact 720

aatgagcaaa aacttcacaa aatagccaat gaacttttgc ttactgaaag agcttatgtc 780

aaccgacttg acctcttaga tcaggtattt tattgcaaac tgttggaaga agcaaaccga 840

ggctcgtttc cagcagagat ggtgaataaa atcttttcta atatttcatc aataaatgcc 900

ttccatagta aattcctctt gccagagctg gagaaacgaa tgcaagaatg ggaaactact 960

cctagaattg gagacatcct tcagaaattg gcaccattcc ttaagatgta tggagaatat 1020

gtgaaaggat ttgataatgc aatggaattg gttaaaaaca tgacagaacg tattccccag 1080

ttcaaatcag tggttgaaga aattcagaaa cagaaaatct gtgggagctt aactttgcag 1140

catcacatgc tagaacctgt tcagcggatt ccccggtatg agatgctcct taaggactat 1200

ctaaggaaat tgcctcctga ttccctggac tggaatgatg ctaaaaaatc acttgaaatt 1260

atatctacag cagcaagcca ttctaatagt gcaataagga aaatggagaa cctaaagaaa 1320

ctcttagaga tttatgaaat gttgggagaa gaagaagaca ttgtaaaccc ttcaaatgaa 1380

ctaataaaag aaggacagat cctcaaacta gctgctcgga acacttcagc acaagaacgc 1440

taccttttct tattcaacaa catgttgctg tactgtgtgc ccaaattcag cttggtaggc 1500

tctaaattca cagttcgaac cagggttggc attgatggaa tgaaaattgt agagactcaa 1560

aatgaagaat atccacatac tttccaggtg tctgggaaag agagaacact ggaactgcag 1620

gccagttctg cgcaagacaa agaagaatgg atcaaggccc ttcaagaaac catcgatgct 1680

tttcatcaaa ggcatgaaac cttcagaaat gcaattgcaa aggataatga cattcactca 1740

gaggtttcta ctgctgagct agggaaaaga gccccaagat ggatccgaga taatgaagtg 1800

acaatgtgta tgaaatgtaa agaacctttc aatgcactga cacgaaggag gcatcattgt 1860

cgagcatgtg gatatgtggt ttgttggaaa tgctccgact acaaagctca acttgaatat 1920

gatggtggta aattgagcaa agtttgtaaa gactgttatc aaatcataag tggattcaca 1980

gacagtgaag aaaaaaaaag aaaaggaatt ttagagattg aatcagcaga agtatctgga 2040

aacagtgtgg tgtgcagctt tcttcagtat atggagaagt caaaaccttg gcagaaagct 2100

tggtgtgtga tccccaagca agaccctctt gtgctgtaca tgtatggtgc cccccaggac 2160

gtcagagccc aggccaccat tccacttctg ggctatgtgg tggatgaaat gccaaggagc 2220

gcagacctgc cacacagttt caaactgacc cagtctaagt ccgtgcacag ctttgctgca 2280

gacagtgagg aactgaagca gaagtggctg aaagtcatcc ttttagctgt cacaggtgag 2340

acaccaggtg gtccaaatga gcatccagcc accttggatg atcatcctga acctaagaaa 2400

aaatcagaat gctgaactcc tccaggacca gccatggtgt ggaggtctca ggacttacag 2460

ctcaagacat tcccagctct tcttacacat ctgctagcac tttatgttga aaaatatagg 2520

cccataaatg caaaaaaaaa aaaaaaaagg gcggccgaaa 2560

<210> 2
<211> 766
<212> PRT
<213> Homo sapiens

<400> 2

Met Glu Glu Ile Lys Pro Ala Ser Ala Ser Cys Val Ser Lys Glu Lys
1 5 10 15

Pro Ser Lys Val Ser Asp Leu Ile Ser Arg Phe Glu Gly Gly Ser Ser

```
          20                  25                  30

Leu Ser Asn Tyr Ser Asp Leu Lys Lys Glu Ser Ala Val Asn Leu Asn
        35                  40                  45

Ala Pro Arg Thr Pro Gly Arg His Gly Leu Thr Thr Thr Pro Gln Gln
    50                  55                  60

Lys Leu Leu Ser Gln His Leu Pro Gln Arg Gln Gly Asn Asp Thr Asp
65                  70                  75                  80

Lys Thr Gln Gly Ala Gln Thr Cys Val Ala Asn Gly Val Met Ala Ala
                85                  90                  95

Gln Asn Gln Met Glu Cys Glu Glu Glu Lys Ala Ala Thr Leu Ser Ser
            100                 105                 110

Asp Thr Ser Ile Gln Ala Ser Glu Pro Leu Leu Asp Thr His Ile Val
        115                 120                 125

Asn Gly Glu Arg Asp Glu Thr Ala Thr Ala Pro Ala Ser Pro Thr Thr
    130                 135                 140

Asp Ser Cys Asp Gly Asn Ala Ser Asp Ser Ser Tyr Arg Thr Pro Gly
145                 150                 155                 160

Ile Gly Pro Val Leu Pro Leu Glu Glu Arg Gly Ala Glu Thr Glu Thr
                165                 170                 175

Lys Val Gln Glu Arg Glu Asn Gly Glu Ser Pro Leu Glu Leu Glu Gln
            180                 185                 190

Leu Asp Gln His His Glu Met Lys Glu Thr Asn Glu Gln Lys Leu His
        195                 200                 205

Lys Ile Ala Asn Glu Leu Leu Leu Thr Glu Arg Ala Tyr Val Asn Arg
    210                 215                 220

Leu Asp Leu Leu Asp Gln Val Phe Tyr Cys Lys Leu Leu Glu Glu Ala
225                 230                 235                 240

Asn Arg Gly Ser Phe Pro Ala Glu Met Val Asn Lys Ile Phe Ser Asn
                245                 250                 255

Ile Ser Ser Ile Asn Ala Phe His Ser Lys Phe Leu Leu Pro Glu Leu
            260                 265                 270

Glu Lys Arg Met Gln Glu Trp Glu Thr Thr Pro Arg Ile Gly Asp Ile
        275                 280                 285

Leu Gln Lys Leu Ala Pro Phe Leu Lys Met Tyr Gly Glu Tyr Val Lys
    290                 295                 300

Gly Phe Asp Asn Ala Met Glu Leu Val Lys Asn Met Thr Glu Arg Ile
305                 310                 315                 320

Pro Gln Phe Lys Ser Val Val Glu Glu Ile Gln Lys Gln Lys Ile Cys
                325                 330                 335

Gly Ser Leu Thr Leu Gln His His Met Leu Glu Pro Val Gln Arg Ile
            340                 345                 350
```

```
Pro Arg Tyr Glu Met Leu Leu Lys Asp Tyr Leu Arg Lys Leu Pro Pro
        355                 360                 365

Asp Ser Leu Asp Trp Asn Asp Ala Lys Lys Ser Leu Glu Ile Ile Ser
    370                 375                 380

Thr Ala Ala Ser His Ser Asn Ser Ala Ile Arg Lys Met Glu Asn Leu
385                 390                 395                 400

Lys Lys Leu Leu Glu Ile Tyr Glu Met Leu Gly Glu Glu Glu Asp Ile
                405                 410                 415

Val Asn Pro Ser Asn Glu Leu Ile Lys Glu Gly Gln Ile Leu Lys Leu
            420                 425                 430

Ala Ala Arg Asn Thr Ser Ala Gln Glu Arg Tyr Leu Phe Leu Phe Asn
        435                 440                 445

Asn Met Leu Leu Tyr Cys Val Pro Lys Phe Ser Leu Val Gly Ser Lys
    450                 455                 460

Phe Thr Val Arg Thr Arg Val Gly Ile Asp Gly Met Lys Ile Val Glu
465                 470                 475                 480

Thr Gln Asn Glu Glu Tyr Pro His Thr Phe Gln Val Ser Gly Lys Glu
                485                 490                 495

Arg Thr Leu Glu Leu Gln Ala Ser Ser Ala Gln Asp Lys Glu Glu Trp
            500                 505                 510

Ile Lys Ala Leu Gln Glu Thr Ile Asp Ala Phe His Gln Arg His Glu
        515                 520                 525

Thr Phe Arg Asn Ala Ile Ala Lys Asp Asn Asp Ile His Ser Glu Val
    530                 535                 540

Ser Thr Ala Glu Leu Gly Lys Arg Ala Pro Arg Trp Ile Arg Asp Asn
545                 550                 555                 560

Glu Val Thr Met Cys Met Lys Cys Lys Glu Pro Phe Asn Ala Leu Thr
                565                 570                 575

Arg Arg Arg His His Cys Arg Ala Cys Gly Tyr Val Val Cys Trp Lys
            580                 585                 590

Cys Ser Asp Tyr Lys Ala Gln Leu Glu Tyr Asp Gly Gly Lys Leu Ser
        595                 600                 605

Lys Val Cys Lys Asp Cys Tyr Gln Ile Ile Ser Gly Phe Thr Asp Ser
    610                 615                 620

Glu Glu Lys Lys Arg Lys Gly Ile Leu Glu Ile Glu Ser Ala Glu Val
625                 630                 635                 640

Ser Gly Asn Ser Val Val Cys Ser Phe Leu Gln Tyr Met Glu Lys Ser
                645                 650                 655

Lys Pro Trp Gln Lys Ala Trp Cys Val Ile Pro Lys Gln Asp Pro Leu
            660                 665                 670
```

```
Val Leu Tyr Met Tyr Gly Ala Pro Gln Asp Val Arg Ala Gln Ala Thr
        675                 680                 685

Ile Pro Leu Leu Gly Tyr Val Val Asp Glu Met Pro Arg Ser Ala Asp
    690                 695                 700

Leu Pro His Ser Phe Lys Leu Thr Gln Ser Lys Ser Val His Ser Phe
705                 710                 715                 720

Ala Ala Asp Ser Glu Glu Leu Lys Gln Lys Trp Leu Lys Val Ile Leu
                725                 730                 735

Leu Ala Val Thr Gly Glu Thr Pro Gly Gly Pro Asn Glu His Pro Ala
            740                 745                 750

Thr Leu Asp Asp His Pro Glu Pro Lys Lys Lys Ser Glu Cys
        755                 760                 765
```

<210> 3
<211> 2298
<212> DNA
<213> Homo sapiens

<400> 3

```
atggaggaaa ttaaacctgc ctctgcttct tgtgtctcaa aagaaaaacc cagtaaggta     60

tcagatctca tcagtcgctt tgaaggaggc agctcattat caaattatag tgatttgaag    120

aaagagtctg ctgtgaacct aaatgctcct agaaccccag gaaggcatgg attgacaacc    180

acacctcaac aaaaactcct ctcccagcac ttgccacaga ggcagggaaa tgatacagat    240

aagactcagg gtgcacagac ttgtgtggcc aacggtgtaa tggcagcaca aaaccagatg    300

gaatgtgagg aggagaaagc tgccactctt agctcagata cttctattca agcttctgaa    360

cccttgcttg atacgcacat agtgaatgga gaaagagatg aaactgccac agctcctgca    420

tcacccacaa cagatagctg tgatggaaat gcttctgaca gtagctacag gactccaggc    480

ataggcccag tgctcccccc agaagaaaga ggggcagaaa cagaaaccaa ggtacaagag    540

agggaaaatg gggaaagccc tctggaactg gagcagctgg accagcacca tgagatgaag    600

gagactaatg agcaaaaact tcacaaaata gccaatgaac ttttgcttac tgaaagagct    660

tatgtcaacc gacttgacct cttagatcag gtattttatt gcaaactgtt ggaagaagca    720

aaccgaggct cgtttccagc agagatggtg aataaaatct tttctaatat ttcatcaata    780

aatgccttcc atagtaaatt cctcttgcca gagctggaga aacgaatgca agaatgggaa    840

actactccta gaattggaga catccttcag aaattggcac cattccttaa gatgtatgga    900

gaatatgtga aaggatttga taatgcaatg gaattggtta aaaacatgac agaacgtatt    960

ccccagttca aatcagtggt tgaagaaatt cagaaacaga aaatctgtgg gagcttaact   1020

ttgcagcatc acatgctaga acctgttcag cggattcccc ggtatgagat gctccttaag   1080

gactatctaa ggaaattgcc tcctgattcc ctggactgga atgatgctaa aaaatcactt   1140
```

gaaattatat ctacagcagc aagccattct aatagtgcaa taaggaaaat ggagaaccta 1200

aagaaactct tagagattta tgaaatgttg ggagaagaag aagacattgt aaacccttca 1260

aatgaactaa taaaagaagg acagatcctc aaactagctg ctcggaacac ttcagcacaa 1320

gaacgctacc ttttcttatt caacaacatg ttgctgtact gtgtgcccaa attcagcttg 1380

gtaggctcta aattcacagt tcgaaccagg gttggcattg atggaatgaa aattgtagag 1440

actcaaaatg aagaatatcc acatactttc caggtgtctg ggaaagagag aacactggaa 1500

ctgcaggcca gttctgcgca agacaaagaa gaatggatca aggcccttca agaaaccatc 1560

gatgcttttc atcaaaggca tgaaaccttc agaaatgcaa ttgcaaagga taatgacatt 1620

cactcagagg tttctactgc tgagctaggg aaaagagccc caagatggat ccgagataat 1680

gaagtgacaa tgtgtatgaa atgtaaagaa cctttcaatg cactgacacg aaggaggcat 1740

cattgtcgag catgtggata tgtggtttgt tggaaatgct ccgactacaa agctcaactt 1800

gaatatgatg gtggtaaatt gagcaaagtt tgtaaagact gttatcaaat cataagtgga 1860

ttcacagaca gtgaagaaaa aaaaagaaaa ggaattttag agattgaatc agcagaagta 1920

tctggaaaca gtgtggtgtg cagctttctt cagtatatgg agaagtcaaa accttggcag 1980

aaagcttggt gtgtgatccc caagcaagac cctcttgtgc tgtacatgta tggtgccccc 2040

caggacgtca gagcccaggc caccattcca cttctgggct atgtggtgga tgaaatgcca 2100

aggagcgcag acctgccaca cagtttcaaa ctgacccagt ctaagtccgt gcacagcttt 2160

gctgcagaca gtgaggaact gaagcagaag tggctgaaag tcatcctttt agctgtcaca 2220

ggtgagacac caggtggtcc aaatgagcat ccagccacct tggatgatca tcctgaacct 2280

aagaaaaaat cagaatgc 2298

<210> 4
<211> 766
<212> PRT
<213> Rattus norvegicus

<400> 4

Met Glu Glu Ser Asn Pro Ala Pro Thr Ser Cys Ala Ser Lys Gly Lys
1 5 10 15

His Ser Lys Val Ser Asp Leu Ile Ser His Phe Glu Gly Gly Ser Val
20 25 30

Leu Ser Ser Tyr Thr Asp Val Gln Lys Asp Ser Thr Met Asn Leu Asn
35 40 45

Ile Pro Gln Thr Pro Arg Gln His Gly Leu Thr Ser Thr Thr Pro Gln
50 55 60

```
Lys Leu Pro Ser His Lys Ser Pro Gln Lys Gln Glu Lys Asp Ser Asp
65                  70                  75                  80

Gln Asn Gln Gly Gln His Gly Cys Leu Ala Asn Gly Val Ala Ala Ala
                85                  90                  95

Gln Ser Gln Met Glu Cys Glu Thr Glu Lys Glu Ala Ala Leu Ser Pro
            100                 105                 110

Glu Thr Asp Thr Gln Thr Ala Ala Ala Ser Pro Asp Ala His Val Leu
        115                 120                 125

Asn Gly Val Arg Asn Glu Thr Thr Thr Asp Ser Ala Ser Ser Val Thr
    130                 135                 140

Asn Ser His Asp Glu Asn Ala Cys Asp Ser Ser Cys Arg Thr Gln Gly
145                 150                 155                 160

Thr Asp Leu Gly Leu Pro Ser Lys Glu Gly Glu Pro Val Ile Glu Ala
                165                 170                 175

Glu Leu Gln Glu Arg Glu Asn Gly Leu Ser Thr Glu Gly Leu Asn Pro
            180                 185                 190

Leu Asp Gln His His Glu Val Lys Glu Thr Asn Glu Gln Lys Leu His
        195                 200                 205

Lys Ile Ala Thr Glu Leu Leu Leu Thr Glu Arg Ala Tyr Val Ser Arg
    210                 215                 220

Leu Asn Leu Leu Asp Gln Val Phe Tyr Cys Lys Leu Leu Glu Glu Ala
225                 230                 235                 240

Asn Arg Gly Ser Phe Pro Ala Glu Met Val Asn Lys Ile Phe Ser Asn
                245                 250                 255

Ile Ser Ser Ile Asn Ala Phe His Ser Lys Phe Leu Leu Pro Glu Leu
            260                 265                 270

Glu Lys Arg Met Gln Glu Trp Glu Thr Thr Pro Arg Ile Gly Asp Ile
        275                 280                 285

Leu Gln Lys Leu Ala Pro Phe Leu Lys Met Tyr Gly Glu Tyr Val Lys
    290                 295                 300

Gly Phe Asp Asn Ala Val Glu Leu Val Lys Asn Met Thr Glu Arg Val
305                 310                 315                 320

Pro Gln Phe Lys Ser Val Thr Glu Glu Ile Gln Lys Gln Lys Ile Cys
                325                 330                 335

Gly Ser Leu Thr Leu Gln His His Met Leu Glu Pro Ile Gln Arg Ile
            340                 345                 350

Pro Arg Tyr Glu Met Leu Leu Lys Asp Tyr Leu Lys Lys Leu Ser Pro
        355                 360                 365

Asp Ala Pro Asp Trp Asn Asp Ala Lys Lys Ser Leu Glu Ile Ile Ser
    370                 375                 380

Thr Ala Ala Ser His Ser Asn Ser Ala Ile Arg Lys Met Glu Asn Leu
```

385 390 395 400

Lys Lys Leu Leu Glu Ile Tyr Glu Met Leu Gly Glu Glu Glu Asp Ile
405 410 415

Val Asn Pro Ser Asn Glu Leu Ile Lys Glu Gly Gln Ile Leu Lys Leu
420 425 430

Ala Ala Arg Asn Thr Ser Ala Gln Glu Arg Tyr Leu Phe Leu Phe Asn
435 440 445

Asn Met Leu Leu Tyr Cys Val Pro Arg Phe Ser Leu Val Gly Ser Lys
450 455 460

Phe Thr Val Arg Thr Arg Val Gly Ile Asp Gly Met Lys Ile Val Glu
465 470 475 480

Thr His Asn Glu Glu Tyr Pro His Thr Phe Gln Val Ser Gly Lys Glu
485 490 495

Arg Thr Leu Glu Leu Gln Ala Ser Ser Glu Gln Asp Lys Glu Glu Trp
500 505 510

Ile Lys Ala Leu Gln Glu Ser Ile Asp Ala Phe His Gln Arg His Glu
515 520 525

Thr Phe Arg Asn Ala Ile Ala Lys Glu Asn Asp Ile Pro Leu Glu Val
530 535 540

Ser Thr Ala Glu Leu Gly Lys Arg Ala Pro Arg Trp Ile Arg Asp Asn
545 550 555 560

Glu Val Thr Met Cys Met Lys Cys Lys Glu Ser Phe Asn Ala Leu Thr
565 570 575

Arg Arg Arg His His Cys Arg Ala Cys Gly His Val Val Cys Trp Lys
580 585 590

Cys Ser Asp Tyr Lys Ala Gln Leu Glu Tyr Asp Gly Gly Arg Leu Asn
595 600 605

Lys Val Cys Lys Asp Cys Tyr Gln Ile Met Ser Gly Phe Ala Glu Ser
610 615 620

Glu Glu Lys Lys Arg Arg Gly Ile Leu Glu Ile Glu Ser Ala Glu Val
625 630 635 640

Ser Gly Asn Ser Glu Val Cys Ser Phe Leu Gln Tyr Met Glu Lys Ser
645 650 655

Lys Pro Trp Gln Lys Ile Trp Cys Val Ile Pro Lys Gln Asp Pro Leu
660 665 670

Val Leu Tyr Met Tyr Gly Ala Pro Gln Asp Val Arg Ala Gln Ala Thr
675 680 685

Ile Pro Leu Leu Gly Tyr Ile Val Asp Asp Met Pro Lys Ser Ala Asp
690 695 700

Leu Pro His Ser Phe Lys Leu Thr Gln Ser Lys Ser Val His Ser Phe
705 710 715 720

```
Ala Ala Asp Ser Glu Glu Leu Lys Gln Lys Trp Leu Lys Ile Ile Leu
                725                 730                 735

Leu Ala Val Thr Gly Glu Thr Pro Asp Gly Pro Ser Glu His Leu Asp
            740                 745                 750

Thr Leu Asp Asn Leu Pro Gly Pro Lys Glu Lys Ser Glu Cys
        755                 760                 765
```

<210> 5
<211> 960
<212> PRT
<213> Mus musculus

<400> 5

```
Met His Gly His Arg Val Pro Gly Gly Pro Gly Pro Ser Asp Pro Glu
1               5                   10                  15

Arg Ser Ala Ala Asn Thr Pro Gly Ala Ala Pro Leu Ala Cys Ala Asp
            20                  25                  30

Ser Asp Pro Gly Ala Leu Glu Pro Gly Leu Pro Val Ser Arg Gly Ser
        35                  40                  45

Gly Thr Ala Leu Gly Gly Pro Leu Asp Pro Gln Phe Val Gly Pro Ser
    50                  55                  60

Asp Ala Ser Leu Gly Ala Pro Pro Ser Ser Arg Val Leu Pro Cys Gly
65                  70                  75                  80

Pro Ser Pro Gln His His Arg Ala Leu Arg Phe Ser Tyr His Leu Glu
                85                  90                  95

Gly Ser Gln Pro Arg Pro Gly Leu His Gln Gly Asn Arg Ile Leu Val
            100                 105                 110

Lys Ser Leu Ser Leu Asp Pro Gly Gln Ser Leu Glu Pro His Pro Glu
        115                 120                 125

Gly Pro Gln Arg Leu Arg Ser Asp Pro Gly Pro Pro Thr Glu Ile Pro
    130                 135                 140

Gly Pro Arg Pro Ser Pro Leu Lys Arg Ala Pro Gly Pro Lys Pro Gln
145                 150                 155                 160

Val Pro Pro Lys Pro Ser Tyr Leu Gln Met Pro Arg Val Leu Pro Pro
                165                 170                 175

Pro Glu Pro Ile Pro Pro Pro Pro Ser Arg Pro Leu Pro Ala Asp Pro
            180                 185                 190

Arg Val Ala Lys Gly Leu Val Pro Arg Ala Glu Ala Ser Thr Ser Ser
        195                 200                 205

Ala Ala Val Ser Ser Leu Ile Glu Lys Phe Glu Arg Glu Pro Val Ile
    210                 215                 220

Val Ala Ser Asp Arg Pro Ala Pro Gly Pro Cys Pro Val Pro Pro Glu
225                 230                 235                 240
```

```
Pro Ala Met Leu Pro Gln Pro Pro Pro Gln Pro Thr Gly Ser Gln Leu
                245                 250                 255

Pro Glu Gly Glu Ala Ser Arg Cys Leu Phe Leu Leu Ala Pro Gly Pro
            260                 265                 270

Arg Asp Gly Glu Lys Val Pro Asn Arg Asp Ser Gly Ile Asp Ser Ile
        275                 280                 285

Ser Ser Pro Ser Asn Ser Glu Glu Thr Cys Phe Val Ser Asp Asp Gly
    290                 295                 300

Pro Pro Ile His Ser Leu Cys Pro Gly Pro Pro Ala Leu Ala Ser Met
305                 310                 315                 320

Pro Val Ala Leu Ala Asp Pro His Arg Pro Gly Ser Gln Glu Val Asp
                325                 330                 335

Ser Asp Leu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Lys Glu Arg
            340                 345                 350

Glu Ile Pro Val Pro Pro Met Glu Arg Gln Glu Ser Val Glu Leu Thr
        355                 360                 365

Val Gln Gln Lys Val Phe His Ile Ala Asn Glu Leu Leu Gln Thr Glu
    370                 375                 380

Lys Ala Tyr Val Ser Arg Leu His Leu Leu Asp Gln Val Phe Cys Ala
385                 390                 395                 400

Arg Leu Leu Glu Glu Ala Arg Asn Arg Ser Ser Phe Pro Ala Asp Val
                405                 410                 415

Val His Gly Ile Phe Ser Asn Ile Cys Ser Ile Tyr Cys Phe His Gln
            420                 425                 430

Gln Phe Leu Leu Pro Glu Leu Glu Lys Arg Met Glu Glu Trp Asp Arg
        435                 440                 445

Tyr Pro Arg Ile Gly Asp Ile Leu Gln Lys Leu Ala Pro Phe Leu Lys
    450                 455                 460

Met Tyr Gly Glu Tyr Val Lys Asn Phe Asp Arg Ala Val Glu Leu Val
465                 470                 475                 480

Asn Thr Trp Thr Glu Arg Ser Thr Gln Phe Lys Val Ile Ile His Glu
                485                 490                 495

Val Gln Lys Glu Glu Ala Cys Arg Asn Leu Thr Leu Gln His His Met
            500                 505                 510

Leu Glu Pro Val Gln Arg Ile Pro Arg Tyr Glu Leu Leu Leu Lys Asp
        515                 520                 525

Tyr Leu Leu Lys Leu Pro His Gly Ser Pro Asp Ser Lys Asp Ala Gln
    530                 535                 540

Lys Ser Leu Glu Leu Ile Ala Thr Ala Ala Glu His Ser Asn Ala Ala
545                 550                 555                 560
```

```
Ile Arg Lys Met Glu Arg Met His Lys Leu Leu Lys Val Tyr Glu Leu
                565                 570                 575

Leu Gly Gly Glu Glu Asp Ile Val Ser Pro Thr Lys Glu Leu Ile Lys
            580                 585                 590

Glu Gly His Ile Leu Lys Leu Ser Ala Lys Asn Gly Thr Thr Gln Asp
        595                 600                 605

Arg Tyr Leu Ile Leu Phe Asn Asp Arg Leu Leu Tyr Cys Val Pro Arg
    610                 615                 620

Leu Arg Leu Leu Gly Gln Lys Phe Thr Val Arg Ala Arg Ile Asp Val
625                 630                 635                 640

Asp Gly Met Glu Leu Lys Glu Ser Ser Asn Leu Asn Met Pro Arg Thr
                645                 650                 655

Phe Leu Val Ser Gly Lys Gln Arg Ser Leu Glu Leu Gln Ala Arg Thr
            660                 665                 670

Glu Glu Glu Lys Lys Asp Trp Val Gln Ala Ile Asn Ser Thr Leu Leu
        675                 680                 685

Lys His Glu Gln Thr Leu Glu Thr Phe Lys Leu Leu Asn Ser Thr Asn
    690                 695                 700

Arg Asp Asp Glu Asp Thr Pro Pro Asn Ser Pro Asn Val Asp Leu Gly
705                 710                 715                 720

Lys Arg Ala Pro Thr Pro Ile Arg Glu Lys Glu Val Thr Met Cys Met
                725                 730                 735

Arg Cys Gln Glu Pro Phe Asn Ser Ile Thr Lys Arg Arg His His Cys
            740                 745                 750

Lys Ala Cys Gly His Val Val Cys Gly Lys Cys Ser Glu Phe Arg Ala
        755                 760                 765

Arg Leu Ile Tyr Asp Asn Asn Arg Ser Asn Arg Val Cys Thr Asp Cys
    770                 775                 780

Tyr Val Ala Leu His Gly Ala Pro Gly Ser Ser Pro Ala Cys Ser Gln
785                 790                 795                 800

His Thr Pro Gln Arg Arg Arg Ser Ile Leu Glu Lys Gln Ala Ser Val
                805                 810                 815

Ala Ala Glu Asn Ser Val Ile Cys Ser Phe Leu His Tyr Met Glu Lys
            820                 825                 830

Gly Gly Lys Gly Trp His Lys Ala Trp Phe Val Val Pro Glu Asn Glu
        835                 840                 845

Pro Leu Val Leu Tyr Ile Tyr Gly Ala Pro Gln Asp Val Lys Ala Gln
    850                 855                 860

Arg Ser Leu Pro Leu Ile Gly Phe Glu Val Gly Pro Pro Glu Ala Gly
865                 870                 875                 880

Glu Arg Pro Asp Arg Arg His Val Phe Lys Ile Thr Gln Ser His Leu
```

885 890 895

Ser Trp Tyr Phe Ser Pro Glu Thr Glu Glu Leu Gln Arg Arg Trp Met
900 905 910

Ala Val Leu Gly Arg Ala Gly Arg Gly Asp Thr Phe Cys Pro Gly Pro
915 920 925

Thr Leu Ser Glu Asp Lys Glu Met Glu Glu Thr Pro Val Ala Ala Ser
930 935 940

Gly Ala Thr Ala Glu Pro Pro Glu Ala Ser Gln Thr Arg Asp Lys Thr
945 950 955 960

<210> 6
<211> 961
<212> PRT
<213> Homo sapiens

<400> 6

Met His Gly His Arg Ala Pro Gly Gly Arg Arg Ala Phe Gly Ala Arg
1 5 10 15

Thr Pro Gly His Glu Pro Ala Gly Ala Ala Pro Pro Ala Cys Ala Asp
20 25 30

Ser Asp Pro Gly Ala Ser Glu Pro Gly Leu Leu Ala Arg Arg Gly Ser
35 40 45

Gly Ser Ala Leu Gly Gly Pro Leu Asp Pro Gln Phe Val Gly Pro Ser
50 55 60

Asp Thr Ser Leu Gly Ala Ala Pro Gly His Arg Val Leu Pro Cys Gly
65 70 75 80

Pro Ser Pro Gln His His Arg Ala Leu Arg Phe Ser Tyr His Leu Glu
85 90 95

Gly Ser Gln Pro Arg Pro Gly Leu His Gln Gly Asn Arg Ile Leu Val
100 105 110

Lys Ser Leu Ser Leu Asp Pro Gly Gln Ser Leu Glu Pro His Pro Glu
115 120 125

Gly Pro Gln Arg Leu Arg Ser Asp Pro Gly Pro Pro Thr Glu Thr Pro
130 135 140

Ser Gln Arg Pro Ser Pro Leu Lys Arg Ala Pro Gly Pro Lys Pro Gln
145 150 155 160

Val Pro Pro Lys Pro Ser Tyr Leu Gln Met Pro Arg Met Pro Pro Pro
165 170 175

Leu Glu Pro Ile Pro Pro Pro Pro Ser Arg Pro Leu Pro Ala Asp Pro
180 185 190

Arg Val Gly Lys Gly Leu Ala Pro Arg Ala Glu Ala Ser Pro Ser Ser
195 200 205

Ala Ala Val Ser Ser Leu Ile Glu Lys Phe Glu Arg Glu Pro Val Ile

```
    210                 215                 220

Val Ala Ser Asp Arg Pro Val Pro Gly Pro Ser Pro Gly Pro Pro Glu
225                 230                 235                 240

Pro Val Met Leu Pro Gln Pro Thr Ser Gln Pro Pro Val Pro Gln Leu
                245                 250                 255

Pro Glu Gly Glu Ala Ser Arg Cys Leu Phe Leu Leu Ala Pro Gly Pro
            260                 265                 270

Arg Asp Gly Glu Lys Val Pro Asn Arg Asp Ser Gly Ile Asp Ser Ile
        275                 280                 285

Ser Ser Pro Ser Asn Ser Glu Glu Thr Cys Phe Val Ser Asp Asp Gly
    290                 295                 300

Pro Pro Ser His Ser Leu Cys Pro Gly Pro Pro Ala Leu Ala Ser Val
305                 310                 315                 320

Pro Val Ala Leu Ala Asp Pro His Arg Pro Gly Ser Gln Glu Val Asp
                325                 330                 335

Ser Asp Leu Glu Glu Glu Asp Asp Glu Glu Glu Glu Glu Glu Lys Asp
            340                 345                 350

Arg Glu Ile Pro Val Pro Leu Met Glu Arg Gln Glu Ser Val Glu Leu
        355                 360                 365

Thr Val Gln Gln Lys Val Phe His Ile Ala Asn Glu Leu Leu Gln Thr
    370                 375                 380

Glu Lys Ala Tyr Val Ser Arg Leu His Leu Leu Asp Gln Val Phe Cys
385                 390                 395                 400

Ala Arg Leu Leu Glu Glu Ala Arg Asn Arg Ser Ser Phe Pro Ala Asp
                405                 410                 415

Val Val His Gly Ile Phe Ser Asn Ile Cys Ser Ile Tyr Cys Phe His
            420                 425                 430

Gln Gln Phe Leu Leu Pro Glu Leu Glu Lys Arg Met Glu Glu Trp Asp
        435                 440                 445

Arg Tyr Pro Arg Ile Gly Asp Ile Leu Gln Lys Leu Ala Pro Phe Leu
    450                 455                 460

Lys Met Tyr Gly Glu Tyr Val Lys Asn Phe Asp Arg Ala Val Glu Leu
465                 470                 475                 480

Val Asn Thr Trp Thr Glu Arg Ser Thr Gln Phe Lys Val Ile Ile His
                485                 490                 495

Glu Val Gln Lys Glu Glu Ala Cys Gly Asn Leu Thr Leu Gln His His
            500                 505                 510

Met Leu Glu Pro Val Gln Arg Ile Pro Arg Tyr Glu Leu Leu Leu Lys
        515                 520                 525

Asp Tyr Leu Leu Lys Leu Pro His Gly Ser Pro Asp Ser Lys Asp Ala
    530                 535                 540
```

```
Gln Lys Ser Leu Glu Leu Ile Ala Thr Ala Ala Glu His Ser Asn Ala
545                 550                 555                 560

Ala Ile Arg Lys Met Glu Arg Met His Lys Leu Leu Lys Val Tyr Glu
                565                 570                 575

Leu Leu Gly Gly Glu Glu Asp Ile Val Ser Pro Thr Lys Glu Leu Ile
            580                 585                 590

Lys Glu Gly His Ile Leu Lys Leu Ser Ala Lys Asn Gly Thr Thr Gln
        595                 600                 605

Asp Arg Tyr Leu Ile Leu Phe Asn Asp Arg Leu Leu Tyr Cys Val Pro
    610                 615                 620

Arg Leu Arg Leu Leu Gly Gln Lys Phe Ser Val Arg Ala Arg Ile Asp
625                 630                 635                 640

Val Asp Gly Met Glu Leu Lys Glu Ser Ser Asn Leu Asn Leu Pro Arg
                645                 650                 655

Thr Phe Leu Val Ser Gly Lys Gln Arg Ser Leu Glu Leu Gln Ala Arg
            660                 665                 670

Thr Glu Glu Glu Lys Lys Asp Trp Val Gln Ala Ile Asn Ser Thr Leu
        675                 680                 685

Leu Lys His Glu Gln Thr Leu Glu Thr Phe Lys Leu Leu Asn Ser Thr
    690                 695                 700

Asn Arg Glu Asp Glu Asp Thr Pro Pro Asn Ser Pro Asn Val Asp Leu
705                 710                 715                 720

Gly Lys Arg Ala Pro Thr Pro Ile Arg Glu Lys Glu Val Thr Met Cys
                725                 730                 735

Met Arg Cys Gln Glu Pro Phe Asn Ser Ile Thr Lys Arg Arg His His
            740                 745                 750

Cys Lys Ala Cys Gly His Val Val Cys Gly Lys Cys Ser Glu Phe Arg
        755                 760                 765

Ala Arg Leu Val Tyr Asp Asn Asn Arg Ser Asn Arg Val Cys Thr Asp
    770                 775                 780

Cys Tyr Val Ala Leu His Gly Val Pro Gly Ser Ser Pro Ala Cys Ser
785                 790                 795                 800

Gln His Thr Pro Gln Arg Arg Arg Ser Ile Leu Glu Lys Gln Ala Ser
                805                 810                 815

Val Ala Ala Glu Asn Ser Val Ile Cys Ser Phe Leu His Tyr Met Glu
            820                 825                 830

Lys Gly Gly Lys Gly Trp His Lys Ala Trp Phe Val Val Pro Glu Asn
        835                 840                 845

Glu Pro Leu Val Leu Tyr Ile Tyr Gly Ala Pro Gln Asp Val Lys Ala
    850                 855                 860
```

```
Gln Arg Ser Leu Pro Leu Ile Gly Phe Glu Val Gly Pro Pro Glu Ala
865                 870                 875                 880

Gly Glu Arg Pro Asp Arg Arg His Val Phe Lys Ile Thr Gln Ser His
                885                 890                 895

Leu Ser Trp Tyr Phe Ser Pro Glu Thr Glu Glu Leu Gln Arg Arg Trp
            900                 905                 910

Met Ala Val Leu Gly Arg Ala Gly Arg Gly Asp Thr Phe Cys Pro Gly
        915                 920                 925

Pro Thr Leu Ser Glu Asp Arg Glu Met Glu Glu Ala Pro Val Ala Ala
    930                 935                 940

Leu Gly Ala Thr Ala Glu Pro Pro Glu Ser Pro Gln Thr Arg Asp Lys
945                 950                 955                 960

Thr
```

<210> 7
<211> 733
<212> PRT
<213> Mus musculus

<400> 7

```
Met Glu Leu Gly Arg Ser Ser Ser Thr Pro Gln Glu Glu Ala Ile Ser
1               5                   10                  15

Pro Leu Gly Val Leu Gly Thr Gly Pro Ser Ser Ser Pro Leu Gly Lys
            20                  25                  30

Leu Gln Ala Leu Pro Ile Gly Pro Gly Ala His Arg Gly Ala His Ser
        35                  40                  45

Ser Ser Ala Pro Ala Gly Asp Ser Ser Thr Arg Glu Pro Ser Gly Ala
    50                  55                  60

Met Lys Ile Pro Asn Arg Asp Ser Gly Ile Asp Ser Pro Ser Ser Ser
65                  70                  75                  80

Val Ala Ser Glu Asn Phe Pro Cys Glu Glu Ser Ser Glu Gly Ser Pro
                85                  90                  95

Ser Pro Ala Ile Leu Gly Leu Pro Ser Glu Thr Ala Ser Asp Ser Arg
            100                 105                 110

Val Pro Gln Asp Asn Pro Gln Glu Glu Glu Asp Ser Gly Val Gly Glu
        115                 120                 125

Glu Pro Asp Pro Lys Val Thr Leu Phe Arg Pro Gln Glu Asp Val Ser
    130                 135                 140

Leu Thr Gln Cys Ser Asp Pro Gln Lys Leu Leu His Ile Ala Gln Glu
145                 150                 155                 160

Leu Leu His Thr Glu Glu Ala Tyr Val Lys Arg Leu His Leu Leu Asp
                165                 170                 175
```

Gln Val Phe Cys Thr Lys Leu Thr Glu Ala Gly Ile Pro Leu Glu Val
180 185 190

Thr Thr Gly Ile Phe Ser Asn Ile Ser Ser Ile Tyr Arg Phe His Gly
195 200 205

Gln Phe Leu Leu Pro Glu Leu Gln Lys Arg Ile Thr Glu Glu Trp Asp
210 215 220

Thr Asn Pro Arg Leu Gly Asp Ile Leu Gln Lys Leu Ala Pro Phe Leu
225 230 235 240

Lys Met Tyr Gly Glu Tyr Val Lys Asn Phe Asp Arg Ala Met Gly Leu
245 250 255

Val Ser Thr Trp Thr Gln Arg Ser Pro Gln Phe Lys Asp Val Ile His
260 265 270

Thr Ile Gln Lys Gln Glu Val Cys Gly Asn Leu Thr Leu Gln His His
275 280 285

Met Leu Glu Pro Val Gln Arg Val Pro Arg Tyr Glu Leu Leu Leu Lys
290 295 300

Asp Tyr Leu Lys Arg Leu Pro Arg Asp Ala Pro Asp Arg Lys Asp Ala
305 310 315 320

Glu Arg Ser Leu Glu Leu Ile Ser Thr Ala Ala Asp His Ser Asn Ala
325 330 335

Ala Ile Arg Lys Met Glu Lys Met His Lys Leu Leu Glu Val Tyr Glu
340 345 350

Gln Leu Gly Gly Glu Glu Asp Ile Val Asn Pro Ala Asn Glu Leu Ile
355 360 365

Lys Glu Gly Ser Ile Gln Lys Leu Ser Ala Lys Asn Gly Thr Thr Gln
370 375 380

Asp Arg His Leu Phe Leu Phe Asn Asn Val Met Leu Tyr Cys Val Pro
385 390 395 400

Lys Leu Arg Leu Met Gly Gln Lys Leu Ser Val Arg Glu Lys Met Asp
405 410 415

Ile Ser Asp Leu Gln Val Gln Asp Ile Val Lys Pro Asn Ala Ala Cys
420 425 430

Thr Phe Ile Ile Thr Gly Arg Lys Arg Ser Leu Glu Leu Gln Thr Arg
435 440 445

Thr Glu Glu Glu Lys Lys Glu Trp Ile Gln Val Ile Gln Ala Thr Val
450 455 460

Glu Lys His Lys Gln Lys Ser Glu Thr Phe Arg Ala Phe Gly Gly Ala
465 470 475 480

Cys Ser Gln Asp Glu Glu Pro Thr Leu Ser Pro Asp Gln Pro Val Met
485 490 495

Ser Thr Ser Ser Val Glu Pro Ala Gly Val Ala Asp Ser Asn Gly Gly

```
            500                 505                 510

Thr Pro Gly Ile Glu Ser Arg Lys Ser Ser Ser Lys Thr Arg Arg Asp
        515                 520                 525

Lys Glu Lys Pro Gly Cys Lys Ser Cys Gly Glu Thr Phe Asn Ser Ile
    530                 535                 540

Thr Lys Arg Arg Tyr Arg Cys Lys Leu Cys Gly Glu Val Ile Cys Arg
545                 550                 555                 560

Lys Cys Ser Glu Phe Lys Ala Glu Asn Ser Lys Gln Ser Arg Val Cys
                565                 570                 575

Arg Glu Cys Phe Leu Glu Glu Pro Leu Val Pro Pro Ser Pro Ser Ser
            580                 585                 590

Glu Thr Pro Thr Glu Leu Lys Gln Asn Ala Glu Lys Pro Pro Ser Val
        595                 600                 605

Asp Pro Arg Pro Ser Leu Leu Cys Gly Thr Leu Asn Leu Ser Asp Asp
    610                 615                 620

Gly Thr Thr Trp Asn Glu Val Trp Ala Ala Ile Pro Glu Ser Asp Pro
625                 630                 635                 640

Gln Val Leu Asp Leu Leu Ala Gly Ser Gln Ala Gly Arg Leu Leu Tyr
                645                 650                 655

Ser Ile Pro Leu Ser Gly Cys Asn Ile Thr Met Pro Asp Pro Glu Glu
            660                 665                 670

Gly Leu Glu Ala Gly Cys Ala Trp Lys Leu His Gln Gly Ser Gln Thr
        675                 680                 685

Trp Trp Leu Ser Ala Pro Ser Thr Lys Leu Gln Gln Cys Trp Leu Lys
    690                 695                 700

Ala Leu Gly Thr Ala Val His Gly Asp Thr Ala Gly Asp Arg Pro Gly
705                 710                 715                 720

Ala Ser Gln Pro Gln Ala Pro Ala Gly Thr Asp Thr Pro
                725                 730

<210>  8
<211>  727
<212>  PRT
<213>  Mus musculus

<400>  8

Met Glu Arg Ala Cys Glu Lys Gln Asp Ser Val Cys Asn Leu Val Ala
1               5                   10                  15

Val Phe Glu Asn Asn Arg Thr Pro Gly Glu Ala Pro Gly Ser His Ser
            20                  25                  30

Leu Glu Asp Gln Pro His Ile Pro Glu His Gln Leu Ser Leu Ser Pro
        35                  40                  45

Glu Pro Trp Glu Ala Pro Pro Val Lys Glu Ala Leu Lys Ser Glu Phe
```

```
    50                  55                  60

Arg Pro Val Ser Arg Thr Tyr Leu Ser Ser Leu Lys Asn Lys Leu Ser
65                  70                  75                  80

Ser Gly Ala Trp Arg Arg Ser Cys Gln Pro Gly Val Ser Pro Gly Pro
                85                  90                  95

Glu Thr Gln Glu Pro Glu Glu Lys Arg Val Val Arg Glu Leu Leu Glu
            100                 105                 110

Thr Glu Gln Ala Tyr Val Ala Arg Leu His Leu Leu Asp Gln Val Phe
        115                 120                 125

Phe Gln Glu Leu Leu Arg Glu Ala Gly Arg Ser Lys Ala Phe Pro Glu
    130                 135                 140

Asp Val Val Lys Leu Ile Phe Ser Asn Ile Ser Ser Ile Tyr Arg Phe
145                 150                 155                 160

His Ala Gln Phe Phe Leu Pro Glu Leu Gln Arg Arg Val Asp Asp Trp
                165                 170                 175

Ala Ala Thr Pro Arg Ile Gly Asp Val Ile Gln Lys Leu Ala Pro Phe
            180                 185                 190

Leu Lys Met Tyr Ser Glu Tyr Val Lys Asn Phe Glu Arg Ala Ala Glu
        195                 200                 205

Leu Leu Ala Thr Trp Met Asp Lys Ser Gln Pro Phe Gln Glu Val Val
    210                 215                 220

Thr Arg Ile Gln Cys Ser Glu Ala Ser Ser Ser Leu Thr Leu Gln His
225                 230                 235                 240

His Met Leu Glu Pro Val Gln Arg Ile Pro Arg Tyr Glu Leu Leu Leu
                245                 250                 255

Lys Glu Tyr Val Gln Lys Leu Pro Ala Gln Ala Pro Asp Leu Glu Asp
            260                 265                 270

Ala Gln Arg Ala Leu Asp Met Ile Phe Ser Ala Ala Gln His Ser Asn
        275                 280                 285

Ala Ala Ile Ala Glu Met Glu Arg Leu Gln Gly Leu Trp Asp Val Tyr
    290                 295                 300

Gln Arg Leu Gly Leu Glu Asp Asp Ile Val Asp Pro Ser Asn Thr Leu
305                 310                 315                 320

Leu Arg Glu Gly Pro Val Leu Lys Ile Ser Phe Arg Arg Ser Asp Pro
                325                 330                 335

Met Glu Arg Tyr Leu Val Leu Phe Asn Asn Met Leu Leu Tyr Cys Val
            340                 345                 350

Pro Arg Val Leu Gln Val Gly Ala Gln Phe Gln Val Arg Thr Arg Ile
        355                 360                 365

Asp Val Ala Gly Met Lys Val Arg Glu Leu Thr Asp Ala Glu Phe Pro
    370                 375                 380
```

```
His Ser Phe Leu Val Ser Gly Lys Gln Arg Thr Leu Glu Leu Gln Ala
385                 390                 395                 400

Arg Ser Arg Asp Glu Met Val Ser Trp Met Gln Ala Cys Gln Ala Ala
                405                 410                 415

Ile Asp Gln Val Glu Lys Arg Ser Glu Thr Phe Lys Ala Ala Val Gln
            420                 425                 430

Gly Pro Gln Gly Asp Thr Gln Glu Pro Lys Pro Gln Val Glu Glu Leu
        435                 440                 445

Gly Leu Arg Ala Pro Gln Trp Val Arg Asp Lys Met Val Thr Met Cys
    450                 455                 460

Met Arg Cys Gln Glu Pro Phe Asn Ala Leu Thr Arg Arg Arg His His
465                 470                 475                 480

Cys Arg Ala Cys Gly Tyr Val Val Cys Ala Lys Cys Ser Asp Tyr Arg
                485                 490                 495

Ala Glu Leu Lys Tyr Asp Ser Asn Arg Pro Asn Arg Val Cys Leu Thr
            500                 505                 510

Cys Tyr Thr Phe Leu Thr Gly Asn Val Leu Pro Gln Gly Lys Glu Asp
        515                 520                 525

Lys Arg Arg Gly Ile Leu Glu Lys Glu Ala Ser Ala Ala Pro Glu Gln
    530                 535                 540

Ser Leu Val Cys Ser Phe Leu Gln Leu Ile Gly Asp Lys Cys Ser Arg
545                 550                 555                 560

Ser Leu Pro Arg Ser Trp Cys Val Ile Pro Arg Asp Asp Pro Leu Val
                565                 570                 575

Leu Tyr Val Tyr Ala Ala Pro Gln Asp Thr Lys Ala His Thr Ser Ile
            580                 585                 590

Pro Leu Leu Gly Tyr Gln Val Ile Ser Gly Pro Gln Gly Thr Ser Gly
        595                 600                 605

Phe Pro Ala Ala Thr Val Arg Arg Ala Val His Leu Gln Gly Arg Val
    610                 615                 620

Cys Gly Ala Ala Gly Pro Leu Gly Asp Ser Tyr Gln Ala Cys Ala Ser
625                 630                 635                 640

Gly Thr Pro Glu Asp Leu Thr Lys Lys Met Cys Leu Thr Glu Pro Ala
                645                 650                 655

Ala Ser Cys Ser Ser Arg Val His Asp Ser Leu Pro Arg Pro Thr Pro
            660                 665                 670

Trp Phe Tyr His Phe Thr Pro Ser Trp Ala Thr Pro Asp Pro Pro Asp
        675                 680                 685

Val Tyr Thr Glu Thr Ile His Pro Asp Ser Val Ser Ser Arg His Arg
    690                 695                 700
```

```
Pro Phe Pro Ser Gly Arg Tyr His Gln Val Ser Gln Leu Ala Gly Glu
705                 710                 715                 720

Gly Ala Asp Ile Pro Gly Ser
                725

<210>  9
<211>  174
<212>  PRT
<213>  Homo sapiens

<400>  9

Ser Ser Leu Ser Asn Tyr Ser Asp Leu Lys Lys Glu Ser Ala Val Asn
1               5                   10                  15

Leu Asn Ala Pro Arg Thr Pro Gly Arg His Gly Leu Thr Thr Thr Pro
            20                  25                  30

Gln Gln Lys Leu Leu Ser Gln His Leu Pro Gln Arg Gln Gly Asn Asp
        35                  40                  45

Thr Asp Lys Thr Gln Gly Ala Gln Thr Cys Val Ala Asn Gly Val Met
    50                  55                  60

Ala Ala Gln Asn Gln Met Glu Cys Glu Glu Glu Lys Ala Ala Thr Leu
65                  70                  75                  80

Ser Ser Asp Thr Ser Ile Gln Ala Ser Glu Pro Leu Leu Asp Thr His
                85                  90                  95

Ile Val Asn Gly Glu Arg Asp Glu Thr Ala Thr Ala Pro Ala Ser Pro
            100                 105                 110

Thr Thr Asp Ser Cys Asp Gly Asn Ala Ser Asp Ser Ser Tyr Arg Thr
        115                 120                 125

Pro Gly Ile Gly Pro Val Leu Pro Leu Glu Glu Arg Gly Ala Glu Thr
    130                 135                 140

Glu Thr Lys Val Gln Glu Arg Glu Asn Gly Glu Ser Pro Leu Glu Leu
145                 150                 155                 160

Glu Gln Leu Asp Gln His His Glu Met Lys Glu Thr Asn Glu
                165                 170
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 March 2002 (14.03.2002)

PCT

(10) International Publication Number
WO 02/020765 A3

(51) International Patent Classification[7]: C07K 14/47, C12N 15/12, 5/10, 15/62, C12Q 1/68, C07K 16/18, G01N 33/50

(21) International Application Number: PCT/US01/28337

(22) International Filing Date:
10 September 2001 (10.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/231,089 8 September 2000 (08.09.2000) US

(71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventor: GLUCKSMANN, Maria, Alexandra; 33 Summit Road, Lexington, MA 02173 (US).

(74) Agents: COLBY, Gary, D. et al.; Akin, Gump, Strauss, Hauer & Feld, L.L.P., Suite 2000, One Commerce Square, 2005 Market Street, Philadelphia, PA 19103-7986 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
6 March 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/020765 A3

(54) Title: 38646, A GUANINE NUCLEOTIDE EXCHANGE FACTOR AND USES THEREFOR

(57) Abstract: The invention provides isolated nucleic acids molecules, designated 28646 nucleic acid molecules, which encode a novel guanine-nucleotide exchange factor. The invention also provides also provides antisense nucleic acid molecules, recombinant expression vectors containing 38646 nucleic acid molecules, host cells into which the expression vectors have been introduced, and non-human transgenic animals in which a 38646 gene has been introduced or disruped. The invention still further provides isolated 38646 proteins, fusion proteins, antigenic peptides and anti-38646 antibodies. Diagnostic methods utilizing compositions of the invention are also provided. 38646 expression and activity can be modulated to affect cell shape, motility, cytoskeleton organization, and intracellular protein and vesicle localization or to affect the tensile strength or integrity of a tissue.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/28337

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/47 C12N15/12 C12N5/10 C12N15/62 C12Q1/68
C07K16/18 G01N33/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EMBL, EPO-Internal, WPI Data, PAJ, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	OBAISHI H. ET AL.: "Frabin, a Novel FGD1-related Actin Filament-binding Protein Capable of Changing Cell Shape and Activating c-Jun N-terminal Kinase" JOURNAL OF BIOLOGICAL CHEMISTRY, THE AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, INC., vol. 273, no. 30, 24 July 1998 (1998-07-24), pages 18697-18700, XP002207012 US the whole document -& JP 11 346775 A (JAPAN SCIENCE & TECH CORP; OBAISHI HIROSHI) 21 December 1999 (1999-12-21) the whole document --- -/--	1-22

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 August 2002	29. 08. 2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Kools, P

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
L	DATABASE EMBL [Online] standard; RNA; ROD; 4854 BP, 28 July 1998 (1998-07-28) OBAISHI H. ET AL.: "Rattus norvegicus actin-filament binding protein Frabin mRNA, complete cds." Database accession no. AF038388 XP002207013 cited in the application Sequence of the corresponding above cited journal publication the whole document ---	1-22
X	PASTERIS N G ET AL: "Isolation, characterization, and mapping of the mouse Fgd3 gene, a new Faciogenital Dysplasia (FGD1;Aarskog Syndrome) gene homologue" GENE, ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, NL, vol. 242, no. 1-2, January 2000 (2000-01), pages 237-247, XP004196524 ISSN: 0378-1119 page 246, column 1, paragraph 2 -column 2, paragraph 2 ---	1-22
A	CERIONE ET AL: "The Dbl family of oncogenes" CURRENT OPINION IN CELL BIOLOGY, CURRENT SCIENCE, LONDON, GB, vol. 8, no. 2, 1 April 1996 (1996-04-01), pages 216-222, XP002089991 ISSN: 0955-0674 cited in the application the whole document ---	1-22
A	CHERFILS J ET AL: "GEFs: structural basis for their activation of small GTP-binding proteins" TIBS TRENDS IN BIOCHEMICAL SCIENCES, ELSEVIER PUBLICATION, CAMBRIDGE, EN, vol. 24, no. 8, 1 August 1999 (1999-08-01), pages 306-311, XP004174250 ISSN: 0968-0004 the whole document ---	1-22
A	WO 98 57990 A (ONYX PHARMA INC) 23 December 1998 (1998-12-23) the whole document --- -/--	19-22

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	DATABASE EMBL [Online] standard; cDNA; 1627 BP, ROSEN CA, ET AL.: "cDNA encoding novel human neoplastic disease associated polypeptide #63" Database accession no. AAS34829 XP002207014 the whole document -& WO 01 055163 A ((HUMA-) HUMAN GENOME SCI INC) 2 August 2001 (2001-08-02) Sequence ID No 73 claim 4 ---	1,3-7, 16-18
P,X	DATABASE EMBL [Online] standard; protein; 465 AA, ROSEN CA, ET AL.: "Novel human neoplastic disease associated polypeptide #63" Database accession no. AAU21630 XP002207015 the whole document -& WO 01 055163 A ((HUMA-) HUMAN GENOME SCI INC) 2 August 2001 (2001-08-02) Sequence ID No 357 claim 11 ---	8,10-15
P,X	DATABASE EMBL [Online] standard; RNA; HUM; 2931 BP, 31 October 2001 (2001-10-31) ISOGAI T. ET AL.: "Homo sapiens cDNA FLJ32732 fis, clone TESTI2001141, highly similar to Rattus norvegicus actin-filament binding protein Frabin mRNA" Database accession no. AK057294 XP002207016 the whole document -----	1,3-7, 16-18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/28337

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: partial 13, 15, 18, 21
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: partial 13, 15, 18, 21

Present claim 13 relates to a method for detecting the presence of the polypeptide of claim 8 by contacting the sample with a compound which is defined by reference to a desirable characteristic or property, namely selectively binding to the polypeptide of claim 8.

Claim 15 relates to a kit comprising such compound.

Claim 18 relates to a kit comprising a compound that selectively hybridizes to a nucleic acid.

Furthermore, claim 21 relates to a method for modulating the activity of the polypeptide of claim 8 by contacting the polypeptide with a compound which is defined by reference to a desirable characteristic or property, namely binding to the polypeptide of claim 8 in such an amount to modulate the activity of the polypeptide.

The claims cover all compounds having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for only a very limited number of compounds. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the compounds by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts relating to the antibody of claim 14, and polynucleotide sequences hybridizing to the nucleic acid of claim 1.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 01/28337

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9857990 A	23-12-1998	AU 7966498 A	04-01-1999
		CN 1268954 T	04-10-2000
		EP 0996638 A2	03-05-2000
		JP 2002511756 T	16-04-2002
		US 6340575 B1	22-01-2002
		WO 9857990 A2	23-12-1998
WO 01055163 A	02-08-2001	AU 2950801 A	07-08-2001
		AU 3095801 A	07-08-2001
		AU 3645901 A	07-08-2001
		AU 3646001 A	07-08-2001
		AU 3646101 A	07-08-2001
		AU 3646201 A	07-08-2001
		AU 3646301 A	07-08-2001
		AU 3646401 A	07-08-2001
		AU 3646501 A	07-08-2001
		AU 3646601 A	07-08-2001
		AU 3794301 A	07-08-2001
		AU 3794401 A	07-08-2001
		AU 3794701 A	07-08-2001
		AU 3794901 A	07-08-2001
		AU 3795001 A	07-08-2001
		AU 3795101 A	07-08-2001
		AU 3795201 A	07-08-2001
		AU 3795301 A	07-08-2001
		AU 3795401 A	07-08-2001
		AU 3795501 A	07-08-2001
		AU 3795701 A	07-08-2001
		AU 3795801 A	07-08-2001
		AU 3972601 A	07-08-2001
		AU 3972701 A	07-08-2001
		AU 3972801 A	07-08-2001
		AU 4140201 A	07-08-2001
		AU 4140301 A	07-08-2001
		AU 4140401 A	07-08-2001
		AU 4140501 A	07-08-2001
		AU 4140601 A	07-08-2001
		AU 4140701 A	07-08-2001
		AU 4140801 A	07-08-2001
		AU 4140901 A	07-08-2001
		AU 4141001 A	07-08-2001
		AU 4141101 A	20-08-2001
		AU 4141201 A	07-08-2001
		AU 4141301 A	07-08-2001
		AU 4141401 A	07-08-2001
		AU 4141501 A	07-08-2001
		AU 4141601 A	07-08-2001
		AU 4141701 A	07-08-2001
		AU 4141801 A	07-08-2001
		AU 4141901 A	07-08-2001
		AU 4313401 A	07-08-2001
		AU 4313501 A	07-08-2001
		AU 4313601 A	07-08-2001
		AU 4313701 A	14-08-2001
		AU 4526201 A	07-08-2001
		AU 4719001 A	07-08-2001
		AU 4719101 A	07-08-2001